

ETHNOSCIENCE: EXAMINING COMMON SENSE

MILENA NUTI

UNIVERSITY COLLEGE LONDON

Thesis submitted in partial fulfilment of the requirements for the degree of Ph.D.
University of London 2003

ABSTRACT

Ethnoscience is a branch of science that studies humans, seeking to understand their modes of interpretation of the world, the diversity of these systems, and their origins.

(Chomsky 2000, p. 90)

In this thesis I trace ideas about naturalistic inquiry into commonsense understanding through Chomsky's work. I argue that the resulting picture significantly illuminates both the nature of 'common sense' and existing interdisciplinary debates surrounding it. Specifically, I claim that progress in investigating the nature of humans' commonsense understanding of psychology (folk psychology) has been hampered by the same kind of methodological dualism which for so long haunted scientific accounts of language. Following Chomsky, I discuss in general how a rationalist inquiry into cognitive domains other than language could proceed by positing 'learning theories' for organisms in given domains, LT (O, D), and attempting to characterize their interaction with experience and the resulting knowledge structures. I further consider how the Language Faculty and a Science Forming Faculty might contribute to our introspective awareness of the understanding of the world delivered to us by core commonsense principles. In particular, taking into account methodological insights gained from the study of language, I put forward a view of a Psychology Faculty and I adapt some ideas from Grice (1975) as a first step towards a positive proposal. Evidence from current developmental and ethological research is presented in support of my position. I thus draw out substantial similarities both between the study of language and the study of commonsense psychology, and between the objects of study in each case. By defusing possible philosophical objections to the comparison, I articulate a detailed defence of my claims.

CONTENTS

ACKNOWLEDGEMENTS..... 4

INTRODUCTION..... 7

1 OVERVIEW: SCIENCE, COMMON SENSE AND ETHNOSCIENCE..... 11

2 METHODOLOGICAL FOUNDATIONS: LANGUAGE AND COMMON SENSE..... 40

3 ON FOLK SCIENCE, EARLY SCIENCE AND THE SCIENCE-FORMING FACULTY..... 71

4 LINGUISTICS AND THE ETHNOSCIENCE OF COMMONSENSE PSYCHOLOGY..... 102

5 LANGUAGE, MINDREADING AND COMMON SENSE..... 157

CONCLUSION..... 205

APPENDIX..... 207

BIBLIOGRAPHY..... 213

ACKNOWLEDGEMENTS

Without the help of a large number of people what follows would not exist. I am grateful to Deirdre Wilson who encouraged me to pursue my interests wherever they took me and reassured me that what I was doing would come together eventually and helped me to articulate it when it finally did; I could not have wished for a more constructive and stimulating combination of freedom, inspiration and support from a supervisor. Discussions with Robyn Carston and Neil Smith, and their comments on drafts of chapters, were invaluable. I further owe Deirdre, Robyn and Neil a huge debt from my first exposure to linguistics; thanks to them and to Rita Manzini, Villy Rouchota, Hans van der Koot, John Harris, Misi Brody, Vlad Zergerac, for equipping me with the knowledge, skills and confidence to undertake this research. I studied for an MA in Philosophy and a PhD in Linguistics thanks to a four year grant from the British Academy Humanities Research Board. From the outset, I was extremely fortunate to be immersed in the vibrant atmosphere cultivated by philosophers in London. It was during this time that the ideas behind this research began to take shape, as I tried to reconcile my linguistic formation with new-found perspectives. I have very much enjoyed and learnt an enormous amount from seminars, lectures and illuminating conversations with Paul Horwich, Mike Martin, Tim Crane, Fraser MacBride, Scott Sturgeon, Sam Guttenplan, Barry C. Smith, Nick Humphrey, David Papineau and Gabriel Segal. I thank them for their patience and generosity; had they not kindly continued to engage with me after I returned to linguistics, the process of writing a thesis that began to look increasingly like philosophy would have been lonely indeed. I further benefitted from a two month visit to Paris, where I was made

to feel most welcome at CREA, now Institut Jean Nicod; thanks in particular to Roberto Casati, Anne-Laure Fayard, Ira Noveck, Dan Sperber and Pierre Jacob.

In the middle of my PhD, political commitments led me to take time out from academic life. I am grateful to Ken Livingstone for inviting me to be part of his first mayoral election campaign, and for the subsequent opportunities he gave me to work for him and then return to my research. I came to understand a great deal in that time about both propaganda and Macchiavellian intelligence; thanks especially to Harry Barlow, Chris Herring, Bruce Beresford, Madeleine Williams, John McDonnell, Michael Seifert, Sean Geoghegan and Helen Shaw.

I could not have written this thesis without the constant presence and encouragement of many other friends. For hours of conversations and discussions which helped to focus my ideas and made hard thinking both enjoyable and sociable, thanks to George Powell, Matthew Nudds, Alessandro Zijno, Corinne Iten, James Wilson, Heidi Maibom, Daniel Friesner, Tom Crowther, Nick Allott, David Adger. For also keeping me together whatever else I was doing, thanks to Giulia Nuti, Costanza Preti, Vanna De Rosas, Abigail Luthin, Scarlet Buckley, Alison Roy, Vincenzo De Cecco, Lorna Macfarlane, Frederik Neuwahl, Duccio Lombardi. Special thanks to Richard Breheny for both more than anyone.

Finally, I am deeply grateful to the people whose instruction and influence while I was growing up laid the foundations for what I have done: many thanks to Vanna Vannini Corsi, Patrizio Paolinelli, Joan Yakkey, John Tattersdill, Kay Galwey, Vinay Chand, Julia Braybrook, Stephen Buckley, Stevie Buckley, Fiora Cavedon, Luciana Bertolino-Zan, Frances Welch and Mario Nuti.

Mistakes, of course, are my own.

For my parents

INTRODUCTION

This thesis is about the study of common sense.

For any area of interest, there are various ways to individuate it. As a starting point, we can think of common sense as the body of untutored ideas people have about the world and how it works. Common sense is often confronted with science. Common sense ideas are not the same as the scientific ideas employed to understand roughly the same phenomena, even where both kinds of ideas are expressed using the same terms. Commonsense understanding can be contrasted with the understanding of the world derived via the methodology of natural science.

For any area of interest, there are various ways to try and investigate it. For instance, by simple reflection or introspection on the matter. Or by turning to art or literature in order to obtain a deeper understanding. Or by conducting a scientific inquiry into the field of interest. Any of these options could be pursued to gain understanding of commonsense understanding.

This thesis is about scientific investigation into common sense. A major theme is that great progress can be made by taking a lead from the scientific investigation into language. Two competing modes of inquiry into language might qualify as being scientific, one rationalist in spirit and the other empiricist. A main difference between the two is that the empiricist approach presupposes that language is just the result of the individual learning the linguistic practices of a speech community. On the rationalist approach, language is a naturally occurring object in the mind/brain and the scientific investigation into language amounts to determining a theory of the properties of that object. This is of course the Chomskyan programme for linguistics.

Specifically, this thesis explores how to apply the Chomskyan programme of naturalistic inquiry to the domain of common sense. Chomsky has suggested that this may be a viable field of study. I follow Chomsky in referring to such an endeavour as 'ethnoscience'. Ethnoscience as a rationalist naturalistic inquiry into the mind is distinguished by two features, as is the study of language. First, it presumes there are naturally occurring mental structures or objects which can only be described at an intentional level, corresponding to the level of knowledge and concepts. These objects or structures play an explanatory role in our account of behaviour, albeit possibly very indirectly. They are also assumed to be species specific and (at least to a large extent) genetically determined. Second, the inquiry follows general scientific method. In particular, it avoids the deployment of commonsense concepts in the study of its domain. Where there are pre-scientific ideas that apply to phenomena related to our area of concern, these ideas at most provide an indication of the range of phenomena we take as the initial object of investigation and perhaps yield terms which may be borrowed and re-defined for scientific purposes. Other methodological precepts of scientific inquiry are taken to apply to naturalistic ethnoscience, such as standards of proof and justification.

There are many areas of common sense. We have commonsense ideas about the physical world, the movement and momentum of bodies; about the biological world, the categories of life forms and relations among them; about the social world and about the motivations and causes of human behaviour. One aspect of this latter area of common sense has to do with commonsense understanding of behaviour in terms of unobserved but causally efficacious mental states – a kind of commonsense psychology.

The focus of this thesis is on how a naturalistic inquiry into commonsense psychology may proceed and what can be learnt from the parallel naturalistic inquiry into language as pioneered by Chomsky. While some psychologists and philosophers acknowledge the Chomskyan programme for language in advocating their approaches

to the study of aspects of commonsense psychology, they do not always follow it. In some respects, even basic conceptual distinctions have yet to be recognised.

Of the variety of ways referred to by which an area of interest such as common sense might be approached, the introspective-reflective mode has been well exercised in the study of commonsense psychology – this being what much standard, analytical philosophy (of mind) is in the business of doing. Again, in parallel with the case of language, a naturalistic inquiry into commonsense psychology will initially overlap with, or possibly trample on, some areas of vital concern for philosophy. And, as with language, naturalistic inquiry into these areas will have to free itself from the methodological dualism imposed and policed by modern analytical philosophy. The second half of chapter 1 considers what methodological dualism amounts to. In the thesis overall, I comment on what place, if any, the claims of classical philosophy of mind have in relation to naturalistic inquiry into commonsense psychology, and we consider, in particular, the extraordinary epistemic priority of reflection and introspection on the categories of commonsense psychology.

It is hoped that this thesis will clear the ground necessary for naturalistic inquiry into commonsense psychology to advance, and to some extent circumscribe what one might expect from such an enterprise and where one might expect its possible limits to lie. Some aspects of commonsense ideas about the physical and biological world are also considered. Finally, an outline of what a Psychology Faculty might look like, based on independent sets of proposals drawn from linguistics, developmental psychology, ethology and philosophy, is offered.

In chapter 1 the idea of ethnoscience is introduced, looking at what naturalistic inquiry into common sense involves and contrasting ethnoscience with philosophy. In chapter 2 we examine the Chomskyan programme for the study of language, highlighting points of methodology and conceptual distinctions which will be useful for the study of areas of common sense, particularly commonsense psychology. Chapter 3 looks at some of Chomsky's less widely discussed ideas about scientific inquiry into the mind and their relation to the study of common sense. Chapter 4

draws parallels between linguistics and ethnoscience in discussion of commonsense psychology. From the preliminary domain of inquiry, we aim to separate out various aspects of what contributes to an area of investigation identified intuitively. A view of 'folk psychology', incorporating proposals from current research in different fields, is developed. Chapter 5 considers further extending the comparison between language and commonsense psychology and argues that the case for disanalogies between the two domains has been overstated. I consider two related questions of philosophical interest and show how the perspective of naturalistic inquiry provides surprising answers.

1

OVERVIEW: SCIENCE, COMMON SENSE, AND ETHNOSCIENCE

1.1	SCIENCE VERSUS COMMON SENSE.....	13
1.2	THE DOMAIN OF INQUIRY OF ETHNOSCIENCE.....	15
1.3	CORE COMMON SENSE AND REFLECTIVE COMMON SENSE.....	17
1.4	SCIENTIFIC METHOD IN ETHNOSCIENCE: AVOIDING COMMON SENSE IN THE STUDY OF COMMON SENSE.....	21
1.5	METHODOLOGICAL DUALISM.....	23
1.5.1	<i>Philosophical priority of the introspective mode of inquiry.....</i>	24
1.5.2	<i>Linguistics, philosophy of language and ethnoscience.....</i>	26
1.6	NATURALISTIC ETHNOSCIENCE AND PHILOSOPHY.....	31
1.7	LIMITS OF NATURALISTIC INQUIRY.....	34
1.7.1	<i>Problems and mysteries.....</i>	36
1.7.2	<i>Hopes and expectations.....</i>	38

1

OVERVIEW: SCIENCE, COMMON SENSE, AND ETHNOSCIENCE

This chapter introduces the notion of common sense at issue in this thesis and contrasts the understanding of the world provided by science with that provided by common sense. These two ways of viewing the world clearly have different purposes and consequences; the relation between the two perspectives is at its most intricate when common sense is itself what is up for scientific investigation. Having considered how the natural sciences proceed with caution when faced with commonsense interpretations of the phenomena they study, I aim to submit the notion of *common sense* to the same constraints of naturalistic inquiry.

This is not an enterprise which is without precedent. To study the mind taking into account the methodological practices adopted standardly in the natural sciences is an approach that has been eloquently argued for by Chomsky, and has been applied fruitfully in the case of language. Yet the consequences of extending this kind of approach to the domain of commonsense understanding have not really been considered, despite widespread recent interest in various aspects of commonsense understanding and significant similarities between language and common sense.

At least as long ago as 1975 Chomsky discussed “attributing to the organism *two* cognitive structures: (i) a system of beliefs and expectations about the nature and

behaviour of objects, and (ii) a system of language. Suppose that [scientist S] calls the first system ‘common sense’ and the second ‘grammar’.” (Chomsky 1975, p. 139, my emphasis). ‘Grammar’ has since received a lot of attention from linguists. ‘Common sense’ has also been much discussed by many people in different fields – sometimes referred to as ‘folk’ knowledge, it is often divided up into ‘folk physics’, ‘folk biology’ and ‘folk psychology’. Generally, however, discussion of commonsense understanding has proceeded without taking into account the methodology that has (since Chomsky) constrained the scientific study of language.

More recent essays, collected in Chomsky (2000), make passing references to “the study of commonsense concepts as a branch of naturalistic inquiry” (2000, p. 90) and introduce in this context the term “ethnoscience”¹. This strand of thought can be traced through Chomsky’s writing. I intend to pursue these clues and consider some of the consequences of the position that emerges.

1.1 SCIENCE VERSUS COMMON SENSE

There are differences between science and common sense. One difference is that what counts as science is relatively uncontroversial; what counts as common sense is debatable. What counts as *good* science can be difficult to agree upon, but most scientists, it seems to me, take themselves to be engaged in discovering the principles that will explain how and why things work. That is, they want to discover how things *actually* work, not just how it may *seem*, to the ordinary human interacting with the world on a daily basis, that they work. In pursuing this aim, the scientist has few qualms about postulating unobservable entities and new theoretical concepts with limited application in everyday experience if these provide the best solution to

¹ *Ethnoscience* is thus the label for the branch of naturalistic inquiry that studies commonsense concepts. The title of this thesis is intended to emphasize this meaning. However, the term ‘ethnoscience’ pre-dates Chomsky’s use of it. I will discuss different ways in which ‘ethnoscience’ could be understood, and how the ‘ethno’ prefix has been applied, as we proceed.

whatever problem she, as a scientist, happens to be grappling with. In introducing these concepts the scientist is not concerned if the solution so devised bears little resemblance to pre-theoretic discourse or understanding.

As science often investigates phenomena that fall outside the realm of human perception and direct experience, this is a tension which does not always make itself felt. Unless we are involved in some scientific enterprise, the structure of microscopic particles, say, or the position of distant planets are not things we can really claim to have an immediate understanding of or intuitions about. However, some of the phenomena studied by fields like physics, biology, chemistry, astronomy, are events which *are* perceived by humans. Of these events, some are generally and pre-theoretically interpreted (by most humans and scientists on their days off) as having certain causes and consequences; often these events and their interpretations are described linguistically; and some such descriptions are in terms which have rough theoretical counterparts in science.

Science in turn sometimes borrows words from natural language vocabulary in order to label theoretical constructs. Reasons for doing this include the desire to avoid cumbersome neologisms and the hope that the choice of word will evoke the technical notion and provide initial help in latching onto it. In such cases, however, it is important to realise that the familiar word is used in a highly specialised way with a meaning which may go beyond how it is normally understood, and failure to recognise the differences will lead to confusion. For it is not uncommon when involved in scientific inquiry to leave behind everyday concepts and commonsense notions, even when the events under investigation are familiar and of concern to us as individuals. After all, there is no reason to expect without evidence that the world should work in exactly the way that humans happen to think and talk about it.

Yet there are aspects of the way in which humans think and talk about their surrounding environment and their interactions with it which appear to be remarkably uniform and stable. While this does not necessarily tell us much about how the world works or how it is divided up, it may very well tell us something about

humans themselves: what they perceive to be constituents of the world, how they take these to behave and interact in producing everyday experiences, and how all of this is represented and discussed.

1.2 THE DOMAIN OF INQUIRY FOR ETHNOSCIENCE: 'COMMON SENSE'

It is striking that there are aspects of our pre-scientific conception of the world which are widely shared by humans all over the world and which develop in similar ways from a very early age, despite differences in experience and ability, and independently of explicit teaching. Examples of the sort of understanding I have in mind include expectations about the movement of physical objects, trajectories and gravitational pull, biological categorisation, physiognomic recognition, interpretations of the interests and intentions of conspecifics, and much more (see for example Carey 1986, 1988; Carey & Spelke 1994; Leslie 1994a; Sperber, Premack & Premack 1996; Medin & Atran 1999). Of course, this intuitive understanding can also be enriched and expanded and be subject to variation affected by cultural elaboration, associated encyclopaedic knowledge and linguistic description. This further information might be thought of as 'common sense' too. And still it is doubtful whether even that would exhaust the range of elaborate knowledge drawn from proverbs and other forms of traditional wisdom and folklore which might be classed as common sense on the everyday understanding of the term. *Common sense*, then, on its common construal, is a far from homogeneous category.

However, this is a relatively normal observation when identifying candidate domains for scientific inquiry. Refining ordinary notions for scientific purposes is only to be expected. As Fodor puts it: "[...] science often starts *in media res*, finding out what it's really about as it goes along, thereby discovering the essences of issues" (2001, part I). Fodor goes on to note that:

[this] way of proceeding implies a kind of inductive risk: the danger that the paradigm cases, reference to which defines the common ground of argument, may turn out not to be paradigms of anything. In particular, they may not all exemplify the same natural kind. If so, then the issues have to be formed some other way, or dropped.

(Fodor 2000, Part I)

This warning is well-taken, and its import will become ever more apparent in the following chapters as discussion proceeds to include details of particular accounts of what might pre-theoretically be classed as ‘common sense’, and evaluation of existing debates about these issues. Similar considerations hold for whatever is described by the prefix ‘folk’.²

While there is a whole host of different things that might initially be included under the label of ‘common sense’, it is unlikely that they will all share a unified explanation. In part this is because we have commonsense knowledge about a variety of topics which may need to be accounted for separately. A more significant point to recognise is that for any one of these topics, the commonsense knowledge we have is not necessarily all of the same kind. In particular, some of it may be universal and come from innate endowment; some from more or less conscious, and more or less pre-determined, elaboration of any such endowment, with the relation between these aspects of common sense being a further question.

Nevertheless, as long as we bear in mind that, pending further examination, ‘common sense’ is being used very much in a preliminary way to pick out a category which on inspection is likely to come apart, there is no reason to doubt the worth or viability of the exercise, and we may well keep going. Following Chomsky, let us refer to the scientific study of common sense as *ethnoscience*.

² I avoid *folk* where possible in what follows because of the variety of its applications (when it does appear below it is intended in the sense of whatever position is under discussion).

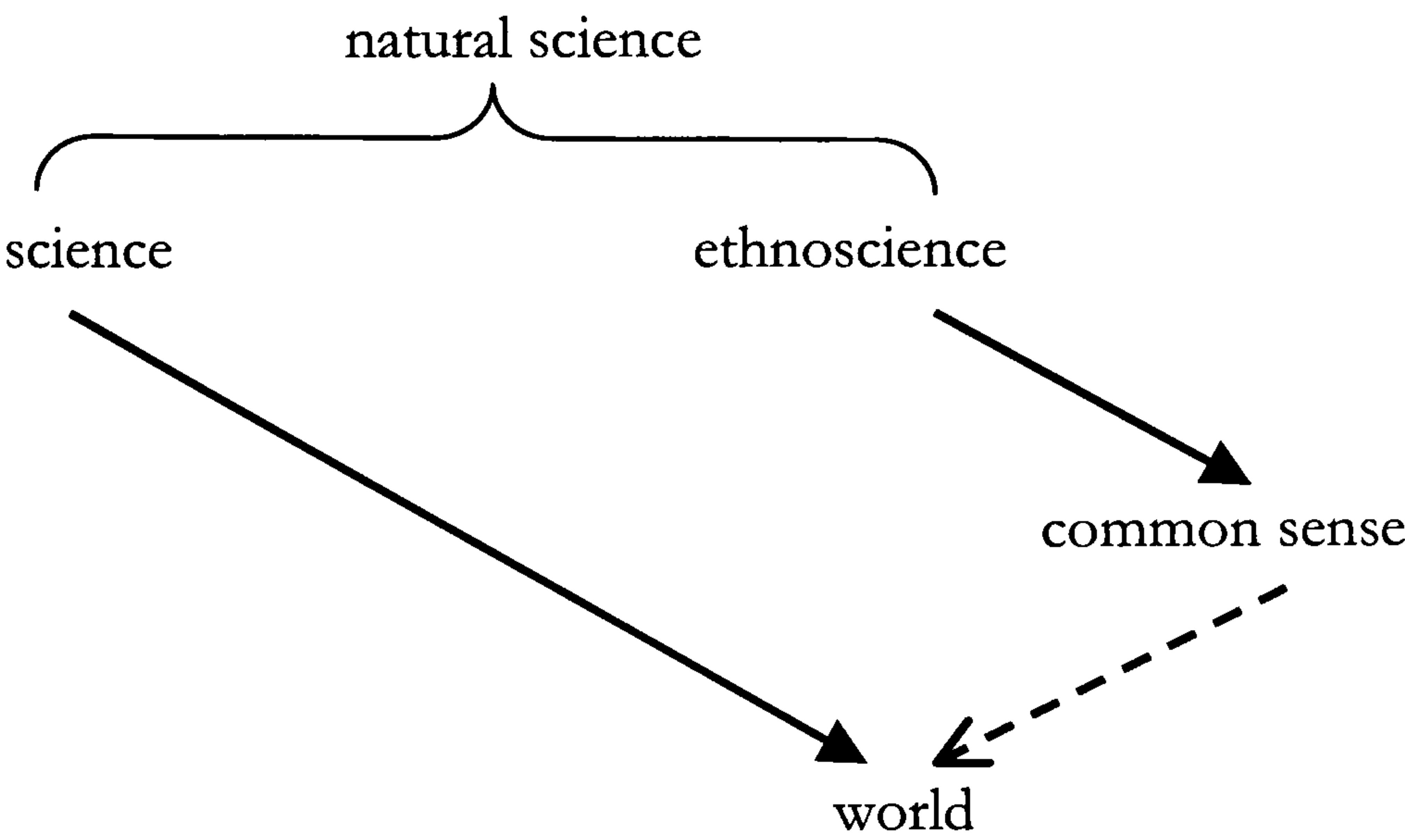
1.3 CORE COMMON SENSE AND REFLECTIVE COMMON SENSE

Ethnoscience is a branch of science that studies humans, seeking to understand their modes of interpretation of the world, the diversity of these systems, and their origins. Separate branches of science study the nature of what humans are sorting out and interpreting in their peculiar ways, whether the phenomena are optical, electrical, mechanical, or mental. Meanwhile, we continue to employ our concepts, sometimes choosing reflectively to refine and modify them, in trying to deal with the problems of ordinary life. These are distinct pursuits.

(Chomsky 2000, p. 90)

Figure (1i) provides an initial illustration of the separate activities at issue (the diagram will be elaborated and significantly modified as we proceed):

Fig. (1i)

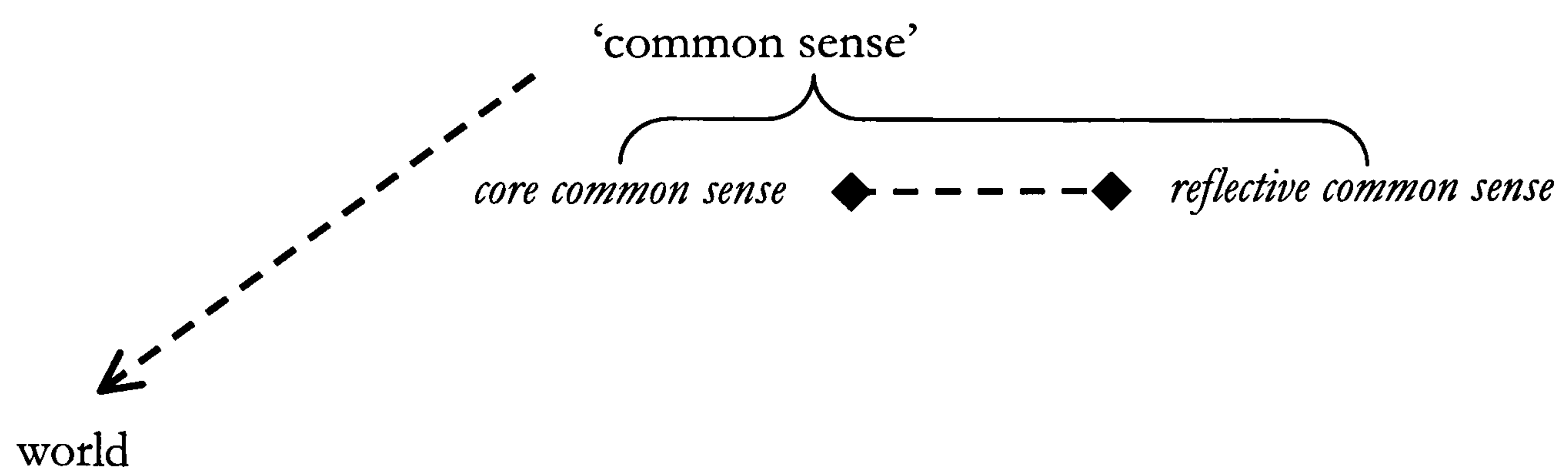


The full arrows indicate scientific inquiry – both natural science and ethnoscience are branches of scientific inquiry. As ethnoscience has as its object of inquiry humans’ understanding of the world, it is part of naturalistic psychology. The dotted arrow indicates humans’ commonsense understanding of some phenomenon. It should be clear that commonsense understanding of some aspect of the world will often not correspond to the best available account provided by natural science. Also it should

be uncontroversial that, for any aspect of the world which finds itself at the focus of attention of both natural science and common sense, the scientific account of commonsense understanding of that aspect of the world, provided by ethnoscience, is not in competition with the natural science explanation of whatever phenomena are under consideration. All of these observations hold whether the aspect of the world at issue is physical or mental.

I think figure (1i) and its discussion are representative of the view expressed in the quote. However, on closer consideration, they do not exhaust what is said. According to figure (1i), the number of “distinct pursuits” illustrated is three: one carried out by humans generally, one by scientists, and one carried out by ethnoscientists in particular. In fact, four are mentioned by Chomsky, as shown in figure (1ii). Aside from the pursuits of scientists and ethnoscientists, humans both (a) understand the world and (b) reflect and talk about how they do so, albeit not always from a scientific perspective. These two activities are shown by the dotted lines in figure (1ii).

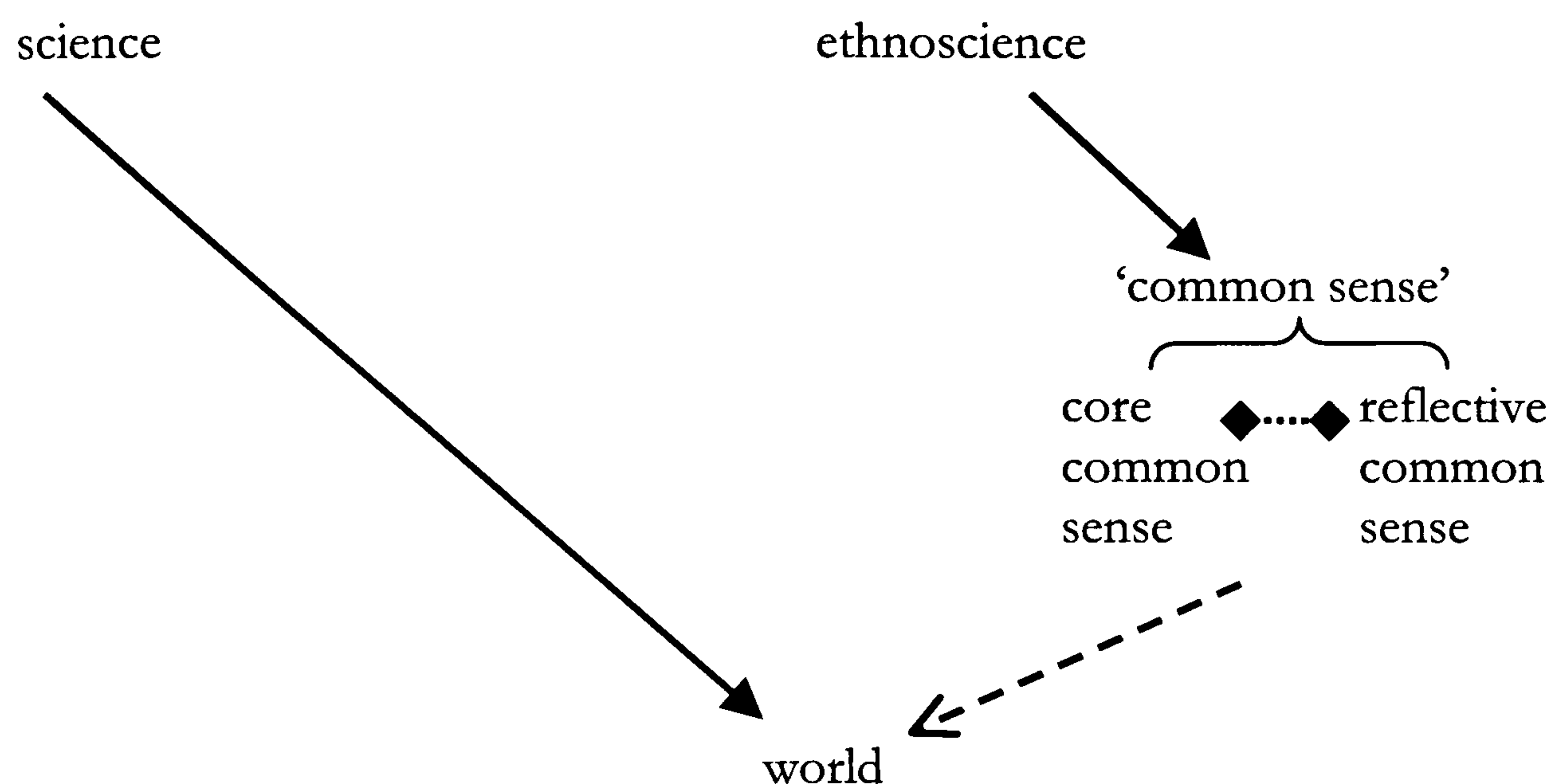
Fig. (1ii)



Here, the dotted arrow pointing towards the world represents (a) “[humans] modes of interpretation of the world”. The connection drawn between core common sense and reflective common sense indicates (b), the way “we [humans] continue to employ our concepts, sometimes choosing reflectively to refine and modify them, in trying to deal with the problems of ordinary life”. The new connection is represented by a

dotted line rather than a full line because I take it that refining and modifying our concepts is an everyday activity not just a scientific enterprise. The connection is not represented as an arrow because the nature of the relation between core common sense and reflective common sense is yet to be considered. Figure (1iii) shows the effect of this change on the overall picture.

Fig. (1iii)



The addition of reflective common sense and its connection to core common sense does not affect the full arrow pointing from science to the world; however, it does raise a question about whether the full arrow pointing from ethnoscience is directed at the whole complex of core common sense plus reflective common sense. Let us assume that it is, and that similarly the dotted arrow directed at the world is pointing from the whole complex. Humans' modes of interpretation of the world are affected by both core common sense and reflective common sense.

In the text surrounding the quote there is some support for these ideas: that there are two aspects to the kinds of activities that we, as humans rather than scientists, engage in, and that both of these fall within the scope of ethnoscience. "Ethnoscience asks how *people interpret and evaluate* what they find around them" (ibid., p. 90), as well as

“*what are the origins* of [people’s] modes of understanding; specifically, what role does innate endowment play” (ibid., p. 91). Also the “folk-sciences” are described as both “instinctively grounded *and* culturally shaped” (ibid. p. 91; all my emphasis); ethnoscience might hope to explore both of these dimensions.³

The point of the addition in figure (1ii) is not just to raise the issue of recognising some of the potential component parts to common sense as it is broadly construed. It also raises an issue about the relation between these components. Returning to the diagram, this is another reason for using a ‘connection’ rather than an arrow: to illustrate what is likely to be a different kind of relation from the ones already introduced. Arguably, *refining* the concepts that you apply to the world, even if you are refining them in order to better apply them, is a different activity from either applying them in the first place or investigating the initial concepts from an independent scientific standpoint.

There is a further element to the discussion not yet addressed by figure (1iii): the contribution made by natural language, in particular by the lexical resources of natural languages. As well as the relation between core common sense and reflective common sense, the relations between core common sense and natural language, and between natural language and reflective common sense, merit careful consideration. These relations will be discussed as they arise with regard to specific areas of common sense.

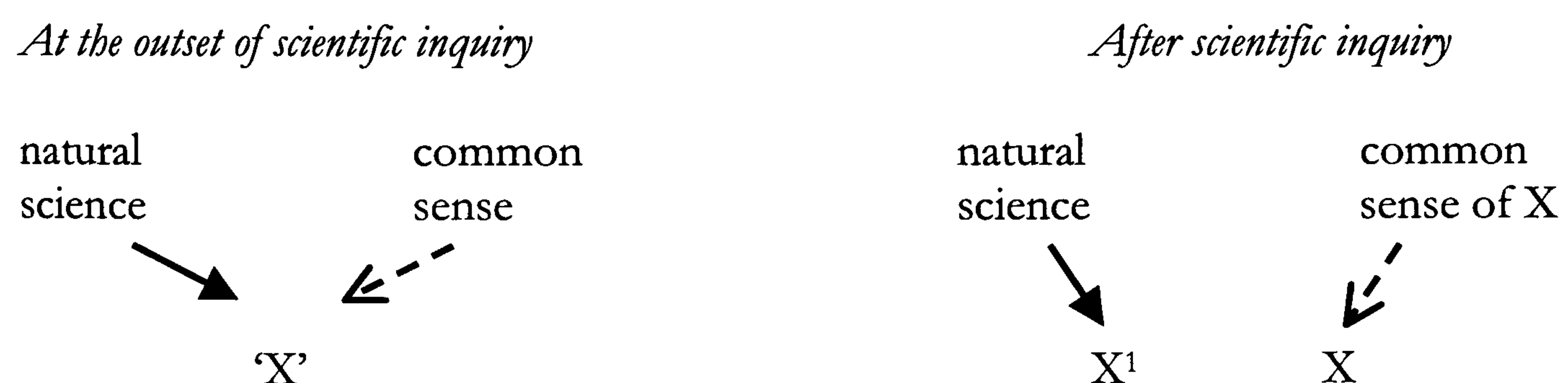
³ Chomsky sometimes uses ‘folk sciences’ to refer to commonsense understanding: “One branch of naturalistic inquiry studies common-sense understanding. Here we are concerned with how people interpret object constancy, the nature and causes of motion, thought and action, and so on (*‘folk science’, in one of the senses of the term*)” (2000, p. 135, my emphasis). Also: “The elements of folk science derive from our biological endowment, taking particular forms under varying cultural conditions.” (ibid., p. 119).

1.4 SCIENTIFIC METHOD IN ETHNOSCIENCE: AVOIDING COMMON SENSE IN THE STUDY OF COMMON SENSE

As far as most branches of the natural sciences are concerned, it is generally accepted that the commonsense perspective on some X should not constrain a scientific account of that X. Even if the intuitive commonsense view is augmented to encompass quite elaborate, reflective ideas and the concepts made available by particular natural languages, these are not usually taken to bear on the status of scientific accounts of X. Furthermore, there is no expectation that a scientific account of X will match up with a scientific account of the commonsense view of X.

In fact it is more than likely, after scientific investigation has taken place, that it will become apparent that science and common sense focus on slightly different entities. Figure (1iv) illustrates a generalised before and after scenario.

Fig. (1iv)



To illustrate these diagrams with an example, take the case of elements recognised since antiquity, earth, wind, fire and water. We might start by looking more closely at what counts, according to the ontology of common sense, as *water*, roughly, the wet stuff in our environment. Of all the different varieties of it, there does appear to be some corresponding natural kind, which can be characterised as having a particular chemical composition. The relatively straightforward case of water contrasts with the cases of the other three in that it is the only one with a clear counterpart in science – not all commonsense categories have one: in as far as commonsense concepts

embody stable assumptions about entities in the world, these assumptions are not necessarily correct.

Even when the ontology of common sense is roughly right, commonsense categories rarely map directly onto scientific ones. Very little of the stuff that we recognise, or describe, as water actually corresponds to H_2O – take acid rain, for instance. Some of what we call ‘water’ is probably less aqueous than some of what we would think of as something else – e.g. a cup of *water* from the Thames as opposed to a cup of *tea* from the pot. And not all of what is H_2O corresponds to what humans readily class as water – pure ice has the same chemical composition.

So, scientific and commonsense concepts tend not to pick out the same things. And the differences between them lie not only in what falls in their extension, but also in the kinds of concepts they are and the purposes they serve. Commonsense concepts are the wrong kind of concepts to appear in a naturalistic theory. Scientific concepts are very precise and unambiguous; they have to be in order to make accurate predictions. The concepts we use in everyday understanding are geared towards other ends; they shift and change and have many variants. Many concepts for everyday purposes are of categories which are *ad hoc* – such as *things to stand on to change a lightbulb* (see Barsalou 1987). Most everyday discourse is used to express concepts of this sort. By contrast, scientific discourse tends to require fixity of meaning, especially where predictions need to be quantifiable.

Commonsense concepts have intricate properties of a sort which are not always simple to elucidate but appear to be known intuitively and are part of what equips them to be used to address the aims and concerns of individuals. In particular, being formed for practical purposes, commonsense concepts typically do not appear to have definable boundaries, a drawback for scientific purposes. Theoretical concepts are constructed so as to avoid this malleability; we understand them differently from how commonsense concepts are understood and we understand the words that express scientific concepts in a different and more

constrained way from how we understand words used to express commonsense concepts.

In short, commonsense concepts are unlikely to identify an object of study that will turn out to be a natural kind; and they are inappropriate tools to feature in a scientific theory of some object of study given the particular complexity of their own structure:

The question is not whether the concepts of commonsense understanding can themselves be studied in some branch of naturalistic inquiry; perhaps they can. Rather, it is whether in studying the natural world (for that matter, in studying these concepts, as part of the natural world), we view it from the standpoint provided by such concepts. Surely not.

(Chomsky 2000, p. 20)

1.5 METHODOLOGICAL DUALISM

The assumptions discussed in 1.4 are widely recognised almost to the point of triviality when the ‘physical’ aspects of the natural world are at issue. However, the situation is less readily accepted in the case of the ‘mental’. When the part of the natural world we are studying is that covered by “Hume’s ‘science of human nature’, which seeks to discover ‘the secret springs and principles, by which the human mind is actuated in its operations’” (Hume quoted in Chomsky 2000, p. 164), new rules suddenly come into play, without these being explicitly acknowledged or justified. It seems philosophy has been the main purveyor of the double standard. As Fodor notes: “To a remarkable extent, and really out of thin air, philosophers have taken it upon themselves to legislate conditions under which empirical inquiry into the mental must proceed.” (Fodor, 2000a). Or, as Chomsky puts it:

Explanatory theories of mind have been proposed, notably in the study of language. They have been seriously challenged, not for violating the canons of methodological naturalism (which they seem to observe, reasonably well), but on other grounds:

‘philosophical grounds,’ which are alleged to show that they are dubious, perhaps outrageous, irrespective of success by the normal criteria of science; or perhaps that they are successful, but do not deal with ‘the mind’ or ‘the mental’ ... [S]uch critiques are commonly a form of methodological dualism, and ... advocacy (or tacit acceptance) of that stance has been a leading theme of much of the most interesting work in recent philosophy of language.

(Chomsky 2000, p. 77)

At best, this kind of dualism results in philosophy and science debating at cross-purposes. However, it can also have more substantive repercussions. As research into the mind becomes more interdisciplinary, to the extent that philosophy hopes to contribute, either by identifying and elucidating the concepts in operation which scientists are supposed to be explaining, or by raising objections to particular theories on the basis of conceptual analysis, it is important that it should share a methodological approach as well as aims and objectives with the sciences over which it sits in judgement.

1.5.1 Philosophical priority of the introspective mode of inquiry

One way in which philosophers have been known to impose extra constraints on inquiry into mental aspects of the world is by holding on tightly to the intuitions about the mind provided by common sense and natural language; much more tightly than would ever seem reasonable even to them in other domains. Contrary to good practice in other areas of naturalistic inquiry, what is taken to be the object of study is frequently determined by reference to notions of everyday concern, without recognition that these are merely a starting point for investigation; and accounts are expected if not required to be cast in terms of recognisable concepts.

For example, when it comes to language, philosophy assumes that the notion of language to be investigated is one which corresponds to (some often complex set of the many) commonsense notions about ‘language’. Having thus identified an arbitrary collection of properties for study, explanations of them tend to be proffered

in terms of familiar concepts; such explanations are then preferred as an account of language and its use over attempts to construct a scientific theory of linguistic ability. There appear to be two main sources of potential confusion here: the first surrounds the object of inquiry and how it is identified; the second stems from the status awarded to explanations that match commonsense intuitions (see also Moravcsik 1975, p. 12-14). So we should be clear whether what is under investigation is language itself or humans' concept of *language*. And any constraint on the kind of conceptual apparatus that is to be allowed to explicate what is under investigation, whether that is language or the concept of *language*, needs to be justified.⁴

Perhaps it would be useful to ask what could make intuitions from natural language and common sense more acceptable as the basis for theories about mental and linguistic aspects of the world than they are considered to be for other areas. It is worth acknowledging that when it comes to mental processes, people often feel that they are able to provide some introspective insight into how their minds are operating; it is this intuition that may be behind appeals in philosophy to 'first-person authority', and arguments for 'personal level' accounts of certain phenomena.⁵ But it is an intuition which imposes strong constraints on what kind of account will pass muster, and what kind of terms an account is allowed to appeal to. Such constraints should be motivated by more than intuition; a better way of proceeding might be to try to explore how this intuition arises, how robust it is, and whether it is justified, thus placing the intuition itself firmly within the realm of what is to be explained rather than treating it as an a priori constraint on explanatory frameworks.

It is also worth acknowledging that we have many notions of linguistic and other mental aspects of the world; some of them are highly complex things like

⁴ See Kornblith (2002, chapter 1) for discussion of a similar contrast between investigating knowledge itself and investigating our concept of *knowledge*.

⁵ Moravcsik (1975, chapter 1) considers further reasons why philosophical concerns may constrain accounts of language in particular; see 2.4.2 below for discussion.

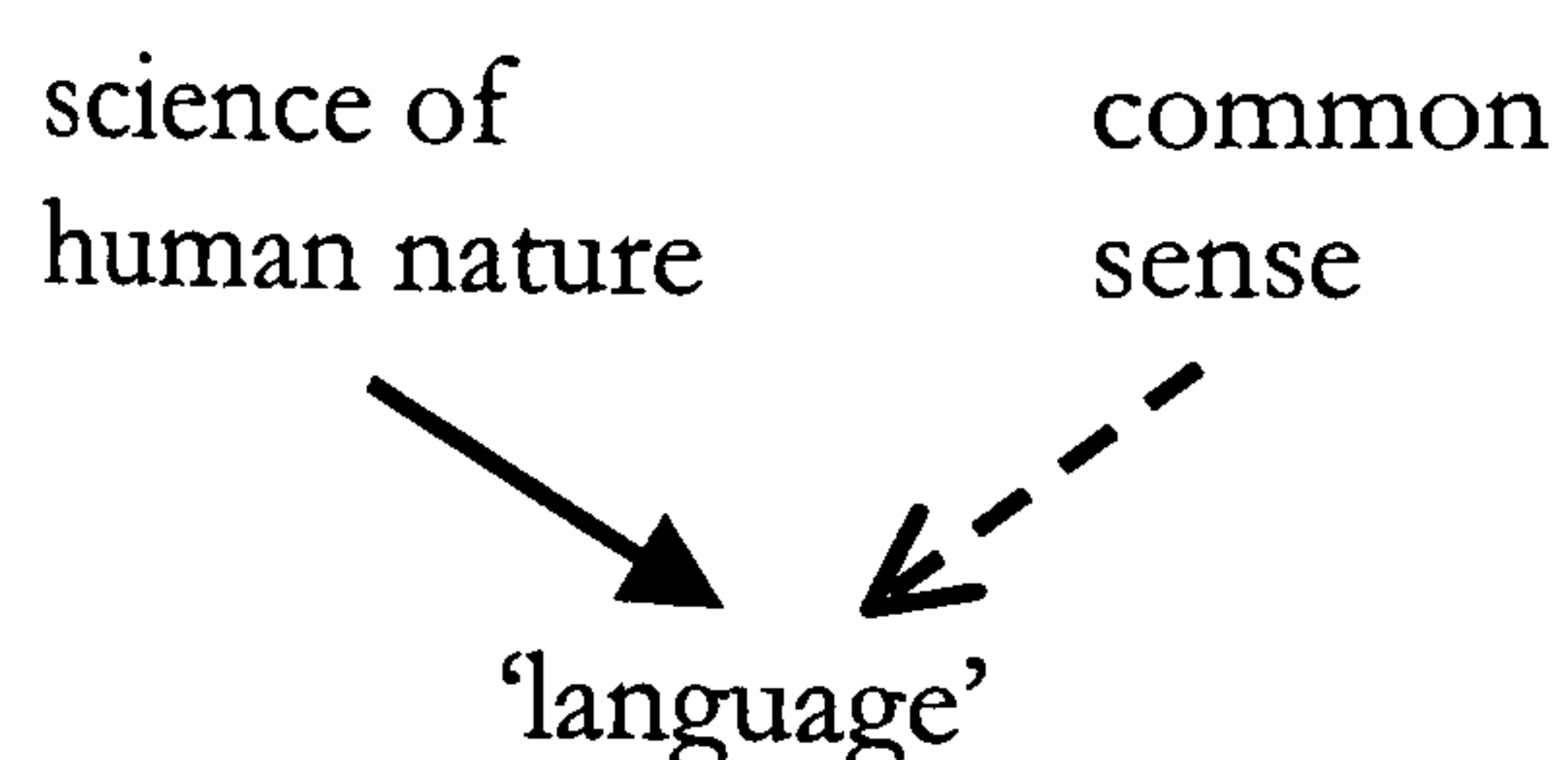
speaking English or *interpreting behaviour*. These are surely of interest and concern to us as individuals, quite probably more so than relatively simpler commonsense concepts like *water*. But if anything, that makes these complex notions even less likely to correspond to natural kinds. While there is little doubt that they are fascinating topics to reflect upon, whether they can be elucidated scientifically as they stand is another matter. These issues will be addressed in 1.7 in discussion of the limits of naturalistic inquiry. First, I would like to elaborate on the repercussions of methodological dualism for the study of language and common sense.

1.5.2 *Linguistics, philosophy of language and ethnoscience*

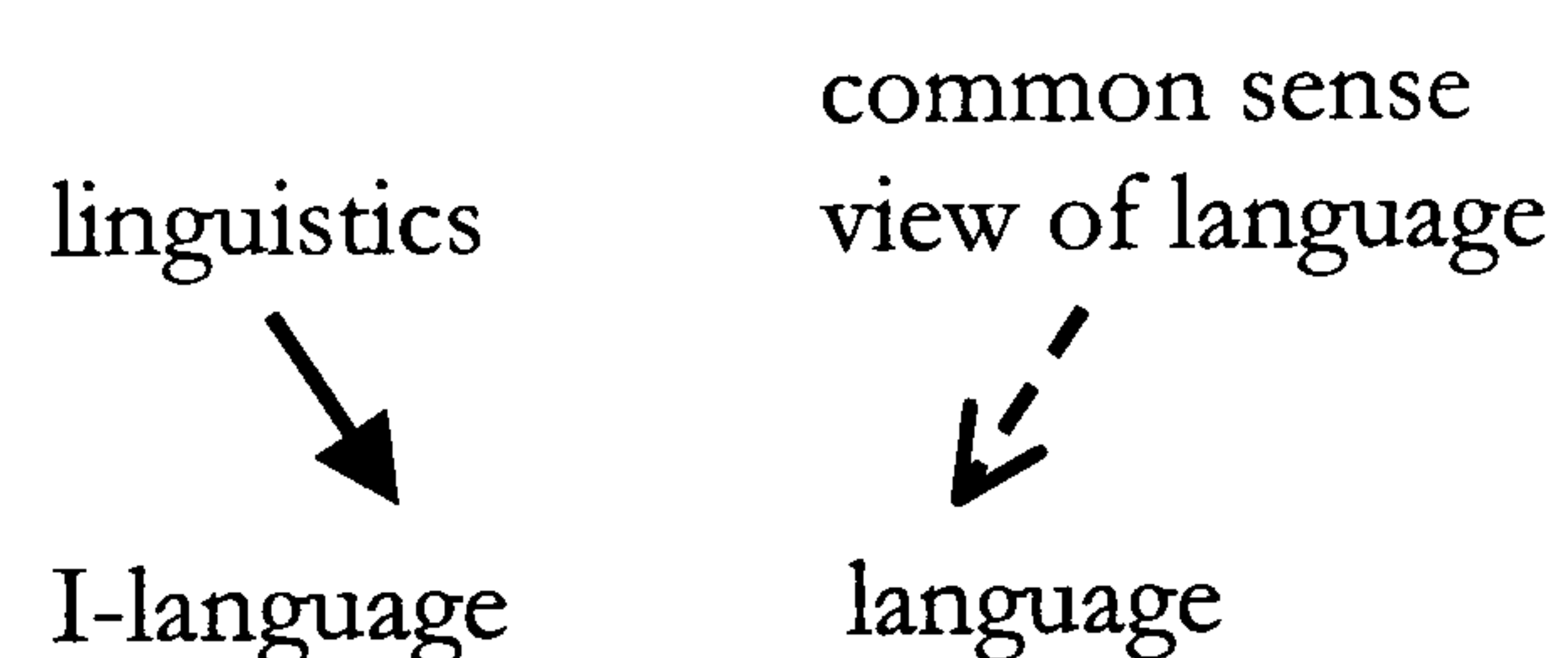
Let us consider potential points of conflict with philosophy if the scientific study of language is faced by tacit acceptance of methodological dualism in philosophical approaches. These observations can then be shown to hold for the study of the mind more generally. To start with, take the object of inquiry. According to methodological naturalism, when it comes to identifying what area is to be studied, the situation for language and mind is really no different from any other branch of natural science. So, slightly modifying the diagram in figure (1iv) we have, roughly:

Fig. (1v)

At the outset of scientific inquiry



After scientific inquiry



Out of all the things that count as language pre-theoretically, science extracts a notion that can be subjected to empirical inquiry (the assumptions which form the basis of this process will be presented in chapter 2). The science in question, linguistics, is part of the science of human nature on the assumption that language is the result of

operations in the brain. Linguistics investigates the language faculty of the mind, and the idea of I-language is as close a counterpart as there is in science to some of the commonsense notions of language.

Of course, I-language is not the only thing that people might wish to investigate. Its study will not answer all the questions that it is possible to ask about language in commonsense terms. But as soon as you take the commonsense concepts of language to be not just your starting point for investigation *but the very thing you want to investigate*, then you have to recognise that you are no longer studying language as a natural object in the mind, but the nature of humans' commonsense conception of 'language'. And if it is an account of the commonsense notions of language that is desired, then we are in the domain of ethnoscience.

Fig. (1vi)

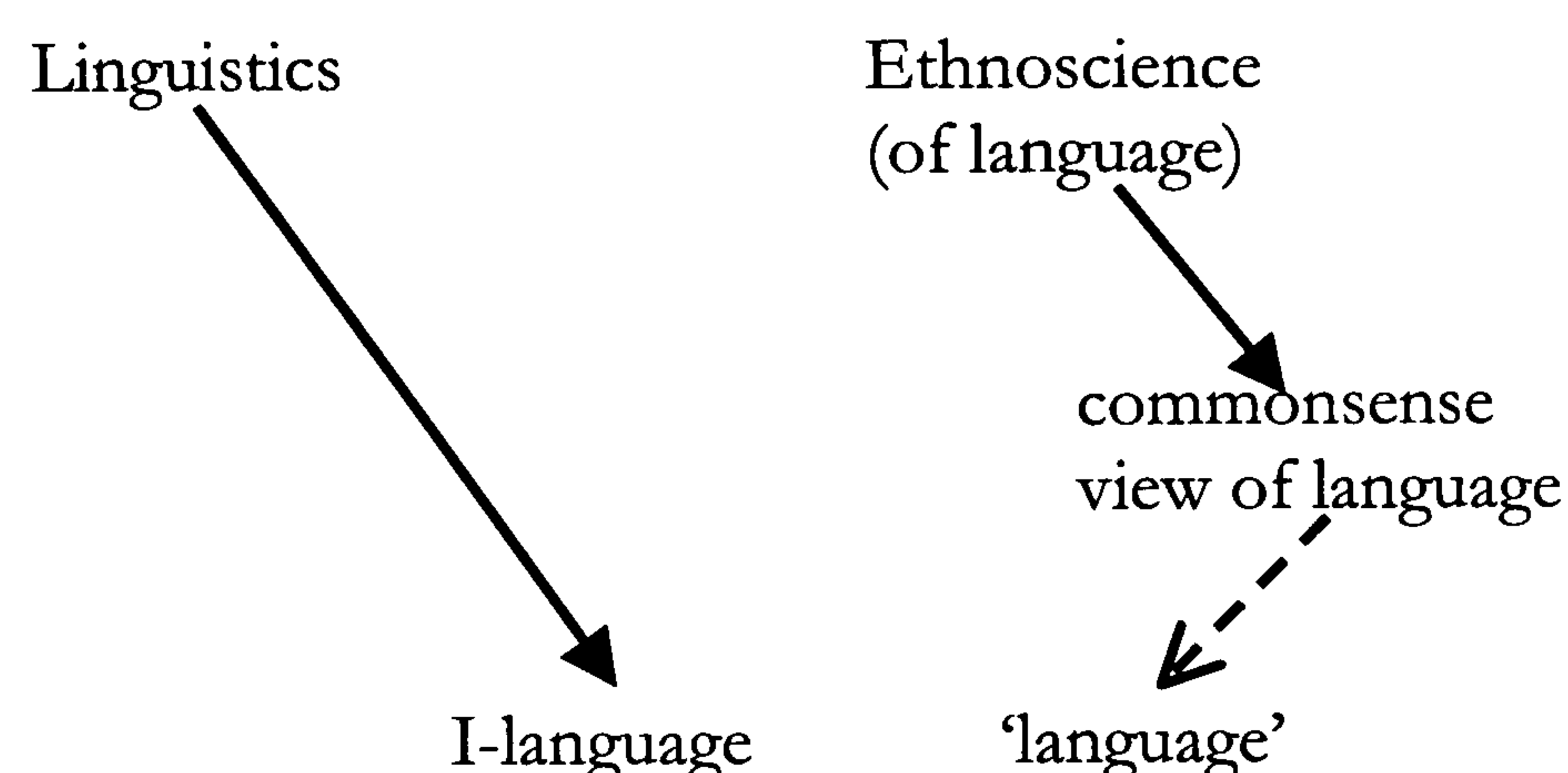


Figure (1vi) is an instantiation of figure (1i) after the process shown in figure (1iv) has taken place. This diagram shows that I-language is a distinct area of inquiry from inquiry into commonsense views of language. And, as argued above (see discussion of figure 1i), commonsense notions about language and accounts of them should not impinge on natural science's explanations of linguistic phenomena. Having established that there are two areas that we might want to investigate – I-language on the one hand, and commonsense views of language on the other – let us turn to the latter, the one covered by ethnoscience (of language). Now, perhaps humans do have systematic commonsense ideas about language and their ability to produce and

understand it. But as the focus of investigation shifts from the language faculty to the conceptual resources humans employ to think about the idea of language pre-scientifically, again familiar methodological questions arise, as we shall see.⁶

If philosophical accounts are mainly concerned with commonsense notions of language, can we say that philosophers are engaged in ethnoscience? Initially, it might seem that the answer to that question is ‘yes’; ethnoscience studies common sense, so insofar as philosophy of language focuses on commonsense notions of language then it could be thought of as ethnoscience of language. Crucially, however, ethnoscience properly construed is part of the science of human nature, and therefore is subject to the constraints of methodological naturalism. For philosophy to count as ethnoscience, it has to avoid falling into the trap of methodological dualism. Let us consider then how philosophy generally approaches the study of (commonsense notions of) language and see whether its work can contribute to ethnoscience.

As argued above, according to methodological naturalism, commonsense concepts are inappropriate tools for scientific investigation, though they can themselves be investigated. Once commonsense notions have been identified as the object of study, scientific explanations of them should not be cast in commonsense terms any more than scientific explanations of other areas are. One of the reasons for not employing commonsense concepts in scientific accounts is to do with the complex and shifting properties of these concepts. Another is to do with the changing ways in which language can be used to express them. This is true of relatively straightforward concepts like *book* and *water* – it is even more significant in the case of the more complex concepts favoured for investigation by philosophers of language and mind, concepts like *truth*, *reference*, *belief*, *meaning*, etc.

Yet it is common practice for philosophy not only to start from natural language when investigating what are, at least initially, pretty clearly commonsense

⁶ Section 5.4 below is devoted to exploring what these ideas are, and whether they constitute a ‘commonsense linguistics’ in any substantive sense.

categories, but also to attempt to elucidate these categories by refining intuitions about the words used to express their meaning; indeed, sometimes it is the explicitly stated aim of an account to provide a semantics for a particular term, as opposed to an account of the commonsense concept underlying it, and occasionally these two aims are equated. But lexical resources do not exhaust conceptual resources – we have more concepts than we do words to express them (Sperber & Wilson 1998). Also, even when we do have a word available to express a concept, given the multifaceted nature of commonsense concepts it is more than likely that the word is picking out just one aspect of the commonsense concept. So the word, even when subjected to detailed analysis, will provide insight into only some part of the concept.

Furthermore, such detailed analysis distances the term from how it is normally understood. Yet judgements about the terms and their application are teased out through thought experiments, experiments which are often strained as intuitions about unfamiliar uses of abstract concepts are tested. Philosophical terms end up being quite technical and are quite far removed from ordinary intuitions. But the terms are technical without being scientific, and are not a very good indication of what the original commonsense concepts amounted to in the first place either.

An alternative approach would allow the possibility that, just as in the case of linguistic ability, where everyday intuitions do not capture what is actually going on (but as data inform theories about linguistic ability), so the conceptual resources underlying commonsense notions about language could go beyond what is available to introspection (with judgements from introspection being part of what is to be accounted for). Of course, there may be connections and similarities between the results of introspection and a scientific explanation, but these are to be discovered rather than assumed.

So it appears that many philosophical approaches tend to be just as accepting of methodological dualism when it comes to the study of commonsense understanding as they are in other aspects of the study of the mind. Although so far for the

purposes of exposition I have concentrated on language and commonsense conceptions of it, these observations are relevant to all of the areas in which ethnoscience has a job to do, wherever humans have commonsense understanding of some aspect of the world and attempts are made to construct a scientific account of exactly what that understanding consists in:

[Issues] are more convoluted when we turn to the science of human nature, which counts among its concerns the investigation of common sense (what we might call ethnoscience). Nevertheless, it proceeds on its own course. Inquiry may begin with ordinary notions of *language, sound, meaning, wind, river*, etc. but without expecting them to be a reliable guide beyond a superficial level.

(Chomsky 2000, p. 164)

If the work done by traditional philosophy cannot be integrated with the project of naturalistic ethnoscience because of the methodological considerations set out above, where in the overall picture does it fit? Perhaps one way of thinking about a particular kind of philosophical approach is to see it as a rarefied form of the reflective common sense sketched in figure (1iii). Taking as a starting point commonsense concepts, philosophers refine intuitions about them, relying on natural language to do so. By contrast, “the ethnoscientist seeks to determine what people take to be constituents of the world, however they may talk about it.” (Chomsky 2000, p. 136).

I expand on the relation between ethnoscience and philosophy immediately below;⁷ subsequent sections deal with foreseeable limits to naturalistic inquiry and map out what ethnoscience might hope to address.

⁷ Implications for philosophy of Chomsky’s approach to the study of the mind have recently been discussed in Bilgrami (2002); Jacob (2002); Stone & Davies (2002).

1.6 NATURALISTIC ETHNOSCIENCE AND PHILOSOPHY

A further clarification will help to avoid the possibility of confusion about what is meant by ‘ethnoscience’ in relation to naturalistic inquiry. It is worth noting that names of sciences can be used to refer to the domain of inquiry of a science as well as the activity of studying it and associated methodology and practices. So if talking about zoology, for example, we could be talking loosely about a set of phenomena or about the discipline that is concerned with them. Syntax is both what syntacticians do, and what they study.

This is a widespread ambiguity which generally goes unnoticed, at least in the case of established sciences; it is either resolved unproblematically, or appears in contexts where it is of no consequence. However, such natural slippage in use is a detail that deserves attention in discussing ethnoscience, as we attempt to demarcate the territory and highlight areas to be investigated. Careful resolution of the ambiguity also bears on the status of philosophical inquiries. To illustrate, note that there are different ways of interpreting a claim that ‘*philosophers do ethnoscience*’ which should be distinguished:

1) Taking ‘ethnoscience’ in its domain-of-inquiry sense, the claim would be that the area of interest to philosophers falls in the domain of ethnoscience. Insofar as philosophers are concerned with the categories humans employ to think about the world pre-theoretically, the statement is unproblematic. To what extent this is in fact true of philosophy could of course be contested, but as far as approaches to language are concerned, it does seem clear, as Chomsky first noted, that most of philosophers’ questions pertain to elucidating concepts – e.g. *language*, *truth* and *reference* – which may play roles in our everyday reasoning *about* language, but are inappropriate for scientific accounts of language itself. For this reason, when a scientific account of language – or rather, I-language – is what is desired, naturalistic approaches to language can be contrasted with the ethnoscience of language: we have different

objects of inquiry and the questions of philosophers belong to ethnoscience. Whether questions of ethnoscience of language are cast naturalistically is then a separate issue.

These points help to explain why, in the context of accounts of language, Chomsky often sounds disparaging about ethnoscience – it is a science of language that is desired by linguistics, not a science of how people think about language pre-theoretically. In other instances – when commonsense notions of language are at issue – ethnoscience, understood as the naturalistic study of commonsense categories, is portrayed as desirable.

2) Taking ‘ethnoscience’ in its label-for-science sense, the claim would become that philosophers are engaged in the discipline of ethnoscience. This statement is about the methodology that is associated with ethnoscience. On the assumption that we are working towards *naturalistic* ethnoscience, what philosophy does is usually not ethnoscience for the reasons I began to outline at the end of the previous section, to do with the assumed theoretical status of commonsense categories and their role in explanations of commonsense understanding. As was noted, philosophy often searches for intuitions about the extension of the terms in the theories it puts forward; but if these terms are theoretical, it is unlikely that we will have reliable intuitions about them.

In linguistics, judgements of native speakers, which form the basis of evidence for many theoretical claims, tap into intuitions that are not directly about the structure of the object of study – subjects are not asked whether they feel an account is a good theory of language or whether it matches their conscious experience, or about the applicability of a theoretical term to an example from a thought experiment. Instead, subjects are asked about a product of this structure – sentences – and the judgements they give are automatic not considered, and not up for debate; indeed to the extent that the judgements are controversial or take a long time to derive, they are of limited, or at least different, value to the theorist. In other words,

subjects are asked *indirectly* about the structure of the theory that is being tested. By contrast:

a good part of contemporary philosophy of language is concerned with analyzing alleged relations between expressions and things, often exploring intuitions about technical notions ‘denote’, ‘refer’, ‘true of’, etc. said to hold between expressions and something else. But there can be no intuitions about these notions, just as there can be none about ‘angular velocity’ or ‘protein’. These are technical terms of philosophical discourse with a stipulated sense that has no counterpart in ordinary language.

(Chomsky 2000, p. 130)

To sum up, ethnoscience as discussed here is what philosophers do in the sense of being what they are interested in doing but it is not what they do in the sense of the activity in which they are engaged. To put it another way, if philosophers are doing ethnoscience, they are doing it badly.

In considering the kinds of questions asked by (mainly externalist) philosophers of language about meaning and reference, Chomsky asks: “Does the inquiry belong to ethnoscience, an investigation of our conceptual resources? *The inquiries that are conducted do not seem to be well designed for this purpose.*” (Chomsky 2000, p. 148). He continues: “*The questions also do not have to do with naturalistic inquiry into the nature of language and its use*, which will develop in its own ways.” (ibid., all my emphasis). This would seem to support the sustainability of the analysis I have presented of the two-fold use of the term ‘ethnoscience’.

3) There is a third possible interpretation of the claim that ‘philosophers do ethnoscience’ which plays more on an ambiguity in ‘do’ than in ‘ethnoscience’. Returning to the domain-of-inquiry sense of the label for a science, that philosophers do ethnoscience *could* mean that what philosophers do is part of the set of phenomena that ethnoscientists (i.e. psychologists investigating commonsense understanding) might want to explain. After all, if philosophy refines concepts from

commonsense to think about language etc., then it could be a particular instantiation of how people think about the world pre-theoretically (or at least non-scientifically). This is the suggestion that was briefly touched upon at the end of the previous section. As it turns out, however, I doubt that the statement expressed by this third interpretation amounts to a tenable position. It is instructive to examine why not.

We have acknowledged that the concepts examined by philosophy, while starting out as those that feature in some part of common sense, are fine-tuned well beyond the scope of ordinary use. This is a highly specialised and inventive activity, which results in the conscious development of technical terms quite far removed from normal discourse or thought. What philosophers are doing seems too unique or idiosyncratic to be part of what ethnoscience can be expected to account for; it is not a process undertaken naturally by most humans, and it is unlikely to be subserved by specialised cognitive structures.

Agreeing with these observations involves implicitly adopting some basis for a preliminary evaluation of a domain as a suitable (or unsuitable) candidate for ethnoscientific explanation. Chapter 2 will elaborate what such criteria may be, building on a model of inquiry familiar from linguistics; we can then return to the question of how much of what could be construed as falling in the domain of ethnoscience is a hopeful or plausible area for some branch of ethnoscience to investigate. Now, having cleared some of the ground for progressing towards a naturalistic approach to ethnoscience, let us set out some realistic expectations.

1.7 LIMITS OF NATURALISTIC INQUIRY

There are two ways in which we might take naturalistic inquiry to have limits; both are worth commenting on. The first concerns the extent to which different phenomena that could be interesting candidates for investigation will turn out to be amenable to this type of inquiry: it is likely that there are limits to what humans can

address naturalistically with any degree of success. The second concerns the extent to which naturalistic inquiry, even when it is successful, provides an exhaustive account of the phenomena it seeks to explain: questions addressed naturalistically may also be considered from other perspectives, with interesting outcomes; indeed, in areas where little naturalistic progress can be made, other kinds of approach will have to be relied upon for understanding.

We could think of these two kinds of limits to naturalistic inquiry as contrasting on the one hand the limits encountered *by humans* in the pursuit of scientific understanding, and, on the other, the limits scientific understanding *itself* is taken to face in providing insight into problems of everyday interest and concern. As this thesis focuses on scientific accounts of the mind, in particular on approaches which conform to the constraints of methodological naturalism as set out above, I will not be addressing the latter issue in detail. Nevertheless, it is important to acknowledge before proceeding that:

Plainly, a naturalistic approach does not exclude other ways of trying to comprehend the world. Someone committed to it can consistently believe (I do) that we learn much more of human interest about how people think and feel and act by reading novels or studying history or the activities of ordinary life than from all of naturalistic psychology, and perhaps always will; similarly, the arts may offer appreciation of the heavens to which astrophysics does not aspire.

(Chomsky 2000, p. 77)

Having established this preliminary, let us now concentrate on the first issue: human limitations in naturalistic inquiry. Of course, if humans are hampered in naturalistic inquiry in some area it could be due to having selected for investigation too disparate phenomena – for example, there could not be a science of ‘objects within a 2m radius of my desk’, as these do not form a natural kind. And it is easy to imagine that more initially plausible candidates might also prove difficult to analyse from a naturalistic perspective for similar reasons – indeed, perhaps even things like ‘understanding

behaviour' and 'language use'. However, this complication is not the main point of the argument here; the important point is rather that even if a natural kind is individuated for investigation, a science of it may fall outside the scope of human capacity.

1.7.1 *Problems and mysteries*

Just as it has been discovered that some tasks are beyond rats, for example, finding the way out of mazes which require numerical concepts for their solution, so it is to be expected, according to Chomsky (e.g. 1993, p. 45), that there are questions which humans are inherently unable to answer. Our cognitive structure restricts the solutions we can come up with – so what actually makes us able to do some things at the same time poses limits on our ability to do others. Questions which lie beyond the grasp of humans' scientific understanding are *mysteries*, as opposed to *problems*. While we may not know beforehand which are which, there is no reason to believe that just because a question can be formulated it will turn out to be a problem. We may well pose questions we are cognitively ill-equipped to answer.

Examples of mysteries often cited by Chomsky are the “creative aspect of language use” and “the causation of behaviour” (e.g. see 1975 pp. 137 onwards; 1979, p. 67 onwards). Part of the reason for Chomsky's scepticism about *scientific* progress in such areas seems to be that the same kinds of questions about them have been asked for thousands of years, always in similar ways, without discernible success. In general he is doubtful that questions relating to what makes people act, react and feel can be answered by humans as scientists, though humans ordinarily have many insights and intuitions on these topics.

However, that the true 'theory of behaviour' is suspected to be beyond humans' cognitive reach does not rule out a theory of humans' commonsense understanding of behaviour. The study of behaviour in the sense of the study of what makes people do the things they do may not have progressed very far, but if people pre-theoretically, systematically, interpret certain behaviours in particular ways,

whether or not they are correct in doing so, this very activity is a potential area of scientific interest. And insofar as it is an activity which is likely to be much more cognitively circumscribed than an all-encompassing ‘theory of behaviour’ would be, it is not an implausible area for study. Indeed, there are passages in Chomsky’s writing that are encouraging on the prospects for a science of commonsense psychological understanding. Bearing in mind that the case of language is the paradigm object of inquiry, consider the following:

Another analogue to the case of language, perhaps, is our comprehension of the social structures in which we live. We have all sorts of tacit and complex knowledge concerning our relations to other people. Perhaps we have a sort of ‘universal grammar’ of possible forms of social interaction, and it is this system which helps us to organize intuitively our imperfect perceptions of social reality.

(Chomsky, 1979, p. 69)

We should then distinguish between the study of what might cause different individuals to exhibit some behaviours rather than others, even in similar circumstances, a probable mystery; and the study of how humans understand and interpret behaviours which are exhibited, a possible problem.

The passage does continue by pointing out that “it does not follow necessarily that we are capable of developing conscious theories in this domain through the exercise of our ‘science-forming faculties’.” (ibid., p. 69-70)⁸ Still, although we cannot take for granted that we are facing an issue that our science-forming faculty, which will be discussed in more detail in chapter 3, is equipped to solve, we can recognise that it is one which does not immediately fall together with mysteries like how choices are made and accounting for free will. If there is some analogue to the case

⁸ Note that the conscious theories at issue should not be understood as conscious theories in this domain that would ever *substitute* for our intuitive understanding, but as conscious theories *of* our intuitive understanding, in the same way that linguists construct conscious theories of linguistic abilities.

of language when it comes to explaining humans' interpretation of psychological matters, trying to ascertain what it is need not encounter the same difficulties as determining what instigated the original behaviours.

The situation is the same in the case of language *use*. Chomsky's doubts about the prospects for a theory of the creative aspect of language might be mis-interpreted as boding ill for the development of pragmatics, or theories of communication. However, while it may not be possible to account for how people are moved to express their thoughts and feelings and in what way, once an utterance has been produced, if humans follow specific procedures to understand how language is used to communicate, elucidating what these procedures exactly amount to again falls within the domain of problem not mystery. This is the difference between 'creative language use' and 'use of language to communicate'. It is not that pragmatics is impossible because all theories of language use are misguided. Rather, just as in the case of behaviour, what is ruled out is a scientific account of certain aspects of production; if humans actually systematically understand (and misunderstand) utterances in predictable ways then there is an interesting line of investigation to be pursued.

1.7.2 *Hopes and expectations*

The reason for all this concern about what falls within the limits of naturalistic inquiry, and what is already expected to lie beyond them, is that it would seem sensible to check before proceeding further that the probable material of ethnoscience is not too ambitious for this type of approach. The preceding discussion sets out some caveats about the slim chances for full 'production' theories of some domains. But in general, there is hope for a science of a domain when we can be clear about the kinds of questions we are asking and have the conceptual resources required to address them. Arguably, this is the position when what is being asked about are the initial and final states of the cognitive structures humans are

endowed with that enable them to build up detailed knowledge about their surrounding environment on the basis of limited and often defective evidence. And there is reason to believe that this line of inquiry may well be worth persevering with beyond the case of language:

Pursuing his efforts to map out the cognitive structures of subjects, S might conclude that each of them possesses an unconscious theory of humans in accordance with which they attribute knowledge of language to other humans. S might also proceed to investigate the physical representation of grammars, theories of humans, common sense, information-processing systems, other systems of factual knowledge and belief, and other cognitive structures that appear to be components of the attained steady states. In this way, he would develop a science of human cognitive structures and, perhaps, their physical basis.

(Chomsky 1975, p. 143)

Of course, it would be possible to agree that there are probably limitations to the extent of human scientific developments but disagree about which areas are likely to exceed these limits and which might be conquered. However, if the items cited in the above quote can be taken as examples of what will be encountered in doing ethnoscience, then we can at least be hopeful about the project, for, imagining ourselves in the place of the scientist S, our “inquiry might take various turns and face innumerable problems, but again, there seems no reason to expect that in this domain [we] would run up against impenetrable mysteries” (ibid., p. 143).

2

METHODOLOGICAL FOUNDATIONS: LANGUAGE AND COMMON SENSE

2.1	LANGUAGE.....	41
2.2	QUESTIONS AND DISTINCTIONS.....	44
2.2.1	<i>Descriptive and explanatory adequacy.....</i>	<i>45</i>
2.2.2	<i>Competence and performance.....</i>	<i>46</i>
2.2.3	<i>Knowledge and mechanisms.....</i>	<i>49</i>
2.2.4	<i>Psychological reality.....</i>	<i>50</i>
2.2.5	<i>Modules.....</i>	<i>52</i>
2.3	LEARNING THEORY FOR ORGANISM O IN DOMAIN D: LT (O, D).....	54
2.3.1	<i>LT (H, L).....</i>	<i>56</i>
2.3.2	<i>LTs for humans in other domains.....</i>	<i>58</i>
2.4	INTRADISCIPLINARY TENSIONS.....	59
2.4.1	<i>Interface issues.....</i>	<i>60</i>
2.4.2	<i>Clashing desiderata.....</i>	<i>68</i>
2.4.3	<i>Looking forward.....</i>	<i>69</i>

2

METHODOLOGICAL FOUNDATIONS: LANGUAGE AND COMMON SENSE

Linguistics studies language and ethnoscience studies common sense. Both disciplines are branches of the science of human nature. I want to use the study of language as a model for the study of common sense. In order to do that, I plan to review the situation in the case of language; then we will be in a better position to see whether the technical notions and the distinctions appealed to in linguistics can be useful in the case of ethnoscience.

2.1 LANGUAGE

Human beings acquire the languages they are exposed to as children; they appear to do so automatically and apparently with little effort. This observation holds true of all (and only) humans, except in cases of specific pathology, and applies to whatever human language any child experiences, up to a certain age. After that, language learning becomes much harder. Objectively, the circumstances under which this task is naturally achieved seem far from ideal. The data available seriously underdetermine the knowledge required to account for the permissible constructions and associated interpretations that everybody takes for granted. Linguistics makes explicit the things humans know about language without instruction – information that is left out of pedagogic grammars as it is assumed without question. For example, it is known

unreflectively that in the sentences '*The men expected to see them*' and '*I wonder who the men expected to see them*', the pronoun '*them*' has to be assigned different referents despite being embedded in similar phrases (see e.g. Chomsky 1986, p. 105). It seems that "to account for the normal use of language we must attribute to the speaker-hearer an intricate system of rules that involve mental operations of a very abstract nature, applying to representations that are quite remote from the physical signal" (Chomsky 1968, p. 53).

Children can be shown to adopt (or develop) structure-dependent rules or principles that operate over abstract representations rather than simpler linear solutions, even without having been exposed to evidence that would justify the more sophisticated strategy. And these rules are not taught – while speakers rely on them they cannot articulate them so would be unable to pass them on. Furthermore, attempts to correct children's speech are made to little avail, and absence of such explicit cues does not impede acquisition or affect the final result. It appears that children do not rely on systematic negative evidence to acquire language. Nor are they put off or confused by ungrammatical strings they are exposed to – people around them are unlikely to speak exclusively in well-formed sentences – despite the fact that these are not marked in any way and are otherwise indistinguishable from acceptable utterances. In short, appealing to experience alone could not explain the judgements about language that humans make intuitively. There is too big a gap between the environmentally available information and the richness of the ensuing cognitive systems. This problem of 'poverty of the stimulus' arises for many areas of cognition, but the case of language has proved perhaps most amenable to study.

A powerful way of explaining these facts about the acquisition of language is to assume that language emerges in virtue of a property of the brain and develops along a largely pre-determined path. As Chomsky frequently puts it, humans grow a language in their mind rather like they grow arms rather than wings. It is this property of the mind/brain that linguists aim to characterise. The focus of scientific

investigation thus shifts from exhibited behaviour to the mental states underlying behaviour. Instead of concentrating on externalised collections of utterances, we concentrate on the internal representations which give rise to them. It is this move from investigating E-language (language conceived of externally, in various ways) to investigating I-languages (internalised systems of knowledge) and their development which leads to linguistics being construed as “part of psychology, ultimately biology” (Chomsky 1986, p. 27).

This stance constitutes a marked departure from previous conceptions of language. It frees the scientific account of language from constraints on it imposed by everyday notions – for example socio-political elements (e.g. similarities and differences between dialects that cut across state boundaries) are not relevant to this investigation, however important their repercussions may be in other areas of human experience. Further departures from common sense have been noted also with regard to other notions that play roles in linguistic theory, among them for example *knowledge, ability, rule, grammar*, as we shall see.

Clearly, there are some differences between human natural languages – Italian and Japanese hardly seem similar, for instance. But differences between languages turn out to be far more superficial than might be assumed at first glance. This fact in turn should be less surprising than it may strike us as being initially. After all, what all natural languages have in common is that they are acquired by humans, so it could be expected that each one should conform to constraints imposed by our cognition. Put briefly, then, the idea is that as part of their cognitive endowment, humans possess a Language Faculty. This innate system, or component of the mind/brain, gives rise to knowledge of language on the basis of triggering experience – it can be thought of as a Language Acquisition Device (LAD). The initial state of the language faculty is common to the species. The genetically determined shared properties of human natural languages are embodied in a Universal Grammar (UG), a set of principles which determine language universals from which particular I-languages can be

derived. An I-language is what is represented in the mind/brain of a speaker who knows a language. Theories of different I-languages derived from UG are given by particular generative grammars.

We can think of the language faculty at various stages of development as being in different states. The initial state, common to everybody, is S_0 . As it is exposed to evidence, the child's language faculty passes through different states until it reaches a steady state S_s . Depending on the evidence available, it will end up determining one or other natural language L . Call the steady state of a particular language S_L . Not everyone who speaks a particular language will have identical I-languages, though the extent to which their I-languages are similar is striking given that even within the same linguistic community people are exposed to very different evidence. Still, there are probably some individual differences; indeed perhaps every one of us to some extent speaks an idiolect. However, whether the similarities between the I-languages of two people with different dialects of French, say, are sufficient for both of them to be counted as French-speakers is a matter of categorisation of what counts as 'French', not a problem that theorists interested in constructing the grammar of either one have to address.

2.2 QUESTIONS AND DISTINCTIONS

The following questions about knowledge of language are characteristic of the approach just outlined (see for example Chomsky 1988, p. 3):

- 1 What knowledge do we have?
- 2 How is this knowledge acquired?
- 3 How is this knowledge put to use?
- 4 What mechanisms implement this knowledge?

I will consider these questions in turn by introducing distinctions that help to address them and which will prove useful in subsequent discussion when we turn from linguistics to ethnoscience. The intention is more to use concepts from the study of language to talk around relevant issues rather than consider in detail the original motivation for their introduction or application. Terminology is of less significance here than the fact that these concepts help us to recognise different areas of interest and to consider different elements in operation in cognitive processes. However, it ought to be noticeable as the discussion proceeds how both technical terms and familiar terms used in their technical sense depart from ordinary discourse and some standard philosophical usages.

2.2.1 *Descriptive and Explanatory Adequacy*

We can relate the first two questions to the ideas of descriptive and explanatory adequacy. The first question – ‘*What knowledge do we have?*’ – is about I-language, about what is represented in the mind of a speaker of a particular language. On the basis of facts about acceptable structures and their permissible interpretations, the linguist’s task is to construct a grammar – a theory of the I-language – that captures these judgements and makes reliable predictions about the status of further strings. A grammar that achieves this can be said to be descriptively adequate.⁹

However, when it comes to constructing grammars of I-languages, there may be more than one solution to the problem. Different grammars might successfully describe the facts. This is the problem of extensionally equivalent theories. To address it, and be able to choose between theories, we need to move beyond just describing the facts and try to explain them, which leads us to the second question. The second question – ‘*How is this knowledge acquired?*’ – is about the initial state of the language faculty and the route that takes the speaker from a universal set of principles to knowledge of a specific language. Answering the second question will involve

⁹ Note that *grammar* in this context (theory of an I-language) has little in common with grammar in the traditional sense of prescriptive rules that are taught and enforced.

constructing a theory of UG. If from the theory of UG it is possible to derive a descriptively adequate theory of an I-language “under the boundary conditions set by experience” (Chomsky 1986, p. 53), then the theory of UG meets the condition of explanatory adequacy, and the grammar it derives can be attributed to the speaker as the theory that is actually in operation, not just a possible theory that might be.

Of course, the task of constructing grammars that accurately describe the facts arises for all attested human natural languages. And remember that the claim is that the principles of UG are the starting point for acquisition of all of these languages. So strong constraints are imposed in two directions: on the one hand, UG has to contain principles that are specific enough to determine rich systems of knowledge on the basis of scant evidence; on the other, the principles have to be general enough to apply to all languages. There is thus a tension between descriptive and explanatory adequacy and at any point data about any language could affect the whole picture. So a study of Italian could well affect a study of Japanese, to take our earlier example. Assumptions about S_0 are implicit in any theory of a S_L . If the assumptions about the initial state entailed by some theory of an I-language for Italian are incompatible with what is known about Japanese, then, given that S_0 must be the same in each case, we will have to adjust the theories. Claims about the universality of linguistic principles are therefore not made lightly and are checked effectively by empirical considerations.

2.2.2 *Competence and performance*

The knowledge that possession of an I-language equips us with (is an essential part of what) enables us to produce and understand novel sentences. This knowledge is *tacit*; it is largely inaccessible to conscious thought processes. This may not match up with some people’s ordinary preconceptions about what knowledge is, which has, on occasion, made it controversial to talk in terms of *knowledge*, a concept which, in

certain philosophical contexts, is identified with justified true *belief*.¹⁰ Knowledge in the case of language therefore is also referred to informally as *competence*. More strictly, *grammatical* competence – which can be equated with I-language – is distinguished from *pragmatic* competence:

Pragmatic competence underlies the ability to use such knowledge along with the conceptual system to achieve certain ends and purposes. It might be that pragmatic competence is characterized by a certain system of constitutive rules represented in the mind, as has been suggested in a number of studies.

(Chomsky 1980, p. 59)

In the context of linguistic theory, competence is contrasted with performance. For example, competence remains constant despite tiredness, drunkenness and physical impediment, but these factors can affect performance. Similarly, performance can be improved while competence remains the same; increased eloquence and rhetorical skill are not necessarily indicators of different levels of competence. Distinguishing between competence and performance helps to avoid confusion in case language-speaking is thought of as an *ability*. Knowledge of language, competence, is an ability that is retained even when the possibility of exercising it is impeded, rather like the ability to ride a bicycle. Performance involves the exercise of abilities.

The third of our questions – ‘*How is this knowledge put to use?*’ – could be understood, then, as a question about performance systems. Indeed, it is often taken in this way and then contrasted with the first question – ‘*What knowledge do we have?*’ – which is taken to be about competence. However, this is an oversimplification. There are two aspects to question 3. Chomsky separates it into what he calls the *perception problem* and the *production problem*. “The perception problem has to do with how we

¹⁰ On *knowledge* and *belief*, see the introduction to the first part of Bernecker & Dretske (2000); Smith (1999, eg. p. 145, p. 161) provides an overview of questions raised by knowledge of language; George (1989) contains philosophical papers on these issues.

interpret what we hear [...] The production problem, which is considerably more obscure, has to do with what we say and why we say it.” (Chomsky 1988, p. 4).¹¹ Both problems relate to how knowledge of language is put to use. Having distinguished between them, it is not clear that either problem matches up straightforwardly with a single notion of performance.

Part of the problem of ‘*how we interpret what we hear*’ might correspond to the notion of pragmatic competence, rather than to a notion of performance; although issues about performance can be raised about pragmatic competence just as they can for grammatical competence. “Theories of grammatical and pragmatic competence must find their place in a theory of performance that takes into account the structure of memory, our mode of organising experience, and so on” (Chomsky 1980, p. 225).

A theory of performance is not just about factors affecting how well a competence is exercised; nor just about the processing systems in virtue of which it is implemented (e.g. a parser). It is also concerned with how a competence is integrated with the rest of the mind. One problem a theory of performance for some competence has to address is how that system interacts with others – which is indeed in some sense still a question about how it is used. Of course, “the study of performance presupposes an understanding of the nature of the cognitive system that is put to use” (Chomsky 1977, p. 49). The task of addressing the nature of a competence is prior to that of attempting to characterise elements to do with performance. As the understanding of grammatical and pragmatic competence develops, associated performance questions become more accessible.

In the absence of such understanding for the problem of *what we say and how we say it*, the contribution of performance factors is even more difficult to assess. Tiredness or anger can cause impatient or terse formulations, even affect whether or

¹¹ This is the same distinction as was introduced here in the argument at the end of section 1.7.1. While serious doubts can be raised about the amenability to scientific study of at least some aspects of the the production problem, the prospects for the perception problem are more hopeful (see also Chomsky 1988, p. 8-9).

y something; fear and adrenaline increase (or decrease) eloquence in certain
i. While “the structure of memory” and “our mode of organising
experience” must play a role in determining *what we say and how we say it* as well as *how
we interpret what we hear*, what that role might be will be much harder to establish in the
former case than in the latter.

2.2.3 *Knowledge and mechanisms*

Linguistics has little to say about how the fourth question relates to knowledge of
language. Investigating ‘*What mechanisms implement this knowledge?*’ is a task that falls to
the brain sciences (and biology more generally). Linguists’ theories of I-languages
and of UG are formulated at a level of abstraction from mechanisms that is justified
by the predictive and explanatory success of their theoretical principles but that is
uncommitted on implementation. The situation is similar to that which applied to
chemistry before it underwent unification with physics.

When we speak of the mind, we are speaking at some level of abstraction of yet-
unknown physical mechanisms of the brain, much as those who spoke of the
valence of oxygen or the benzene ring were speaking at some level of abstraction
about physical mechanisms, then unknown.

(Chomsky 1988, p. 7)

At present, not enough is known about the brain to expect any imminent reduction;
but the principles formulated by linguists can inform us as to what to look for.
“Insofar as the linguist can provide answers to questions 1, 2, and 3, the brain
scientist can begin to explore the physical mechanisms that exhibit the properties
revealed in the linguist’s abstract theory” (ibid, p. 6). Conversely, linguistic theories
eventually might be constrained by discoveries in the brain sciences; faced with
extensionally equivalent theories of an I-language or of UG, a choice between them
could be facilitated if one featured principles easily attributed to established brain
processes while others could not be explained so readily (see Chomsky 1986, p. 39).

the position with regard to the fourth question seems stated quite clearly. Yet there has been some confusion about this issue, with talk of a language acquisition *device*, for example, or of mental *structures*, being understood in physical terms (see e.g. Cowie 2001). Apart from the fact that, given the textual evidence presented above, such a construal appears to be unjustified and not what is intended, there are further reasons to object to it, as pointed out in some detail by Fodor (2001). Briefly, I-languages must comply with principles of UG, so descriptive adequacy of a grammar is not a sufficient condition for it to be true. It is thus an important feature of the approach we have been considering that theories of I-languages are theories about the objects of propositional attitudes rather than about *mechanisms*. Grammars are mentally represented and speaker/hearers stand in the relation of cognizing to them (where *cognizing* is an attitude similar to *knowing* but that is unaffected by the knowledge in question being tacit, see e.g. Chomsky 1980, p. 69). The fact that a grammar describes the facts does not entail that it is what is internally represented, for “representations can differ even if their intentional contents do not” (Fodor 2001). But if the theoretical claims were about *mechanisms* we could not distinguish between a theory that is true and a theory that is just descriptively adequate in this way: “internal mechanisms *aren’t* (normally) internally *represented*; they’re just *internal tout court*. A fortiori, you can’t choose between equivalent theories of an internal mechanism by reference to *how* it is internally represented.” (Fodor, 2001, original emphasis). Nativism about language does not turn on innateness of specific mechanisms for language but on innateness of language-specific ideas, in the rationalist tradition. The language faculty yields a system of knowledge rather than just a processing system.

2.2.4 *Psychological reality*

Normally, in searching for explanations of empirical facts, whether about the structure of the sun, the human visual system, or language, it would be reasonable for scientists, while aware that their theories may well be wrong in part at least, to

by assuming them to be tentatively true and continue research by adding them with more evidence, refining or rejecting them accordingly, hoping to find support for them at other levels of description. The evidence that is taken into account may come from practically anywhere and is not ranked. However, when it comes to the study of language, certain types of evidence have been claimed to be inadmissible; and certain types of evidence have been claimed to be more significant than others.

For example, despite purporting to provide an account of human language, some theories claim not to be ‘psychological hypotheses’ and thus ignore whole categories of evidence (see e.g. Dummett 1976, p. 70-71; Davidson 1984). In order to distinguish between this kind of approach and any approach to some aspect of the mind that does consider itself to be a psychological hypothesis – an actual statement about the mind/brain of a subject, affected and constrained by all types of evidence including that of a psychological nature – we could describe the latter kind of approach as aspiring to *psychological reality*.

However we should be aware, if this is the terminology that we choose to adopt, that the label does no more work than make explicit that distinction. It is one thing to say that a theory has psychological reality as a way of acknowledging that evidence from psychology is accommodated in the theory – a natural constraint for linguists and psychologists but not so uncontroversial for some philosophers. But that is not to say that evidence tagged as psychological is more valuable than, say, linguistic evidence, or that a theory based on facts about language somehow falls short of counting as a psychological theory in the absence of so-called psychological evidence (in fact, at the moment the opposite is more likely to be the case, as e.g. processing evidence is rarely fine-grained enough and often inconclusive). The point is that to draw any such distinction among types of evidence would amount to abandoning without justification the standard realist stance towards postulated theoretical entities. The principles and features of a linguistic theory are intended to be construed as true at some significant level of description. We might ask, then,

‘psychological reality’, as distinct from ‘truth, in a certain domain?’” (Fodor 1980, p. 107). After all, no additional constraint of ‘physical reality’ is required of theories about non-mental aspects of the world.

A parallel issue arises when evolution is brought into the picture. Currently much is made of accounts of mind having ‘evolutionary plausibility’, even though it could be argued that to talk of the evolution of capacities before much is known about what the capacities actually consist in is premature (see chapter 3). The label is widely used as if it provides further validation for the account it describes. But note that facts about evolution are just another category of evidence that needs to be taken into consideration in testing theories – that such facts *are* taken into consideration by some account may make it a better evidenced account, to the extent that it can accommodate them, but does not bestow upon it some further property of ‘evolutionary reality’. Once we have established that we are doing science, describing an account as psychologically real or evolutionarily plausible runs the risk of being vacuous.

2.2.5 *Modules*

The language faculty is described by Chomsky as *modular*:

I am tentatively assuming the mind to be modular in structure, a system of interacting subsystems that have their own special properties. [...] I have been attempting to isolate for special study one of these systems, the faculty of human language.

(Chomsky 1980, p. 89-90)

Modularity is a widespread notion in cognitive science, adopted and adapted by theorists of many different persuasions.¹² As claims about modules in one sense or

¹² Segal (1996) distinguishes kinds of modularity. See also Garfield (1987); Karmiloff-Smith (1992); Sperber (1994a) for different notions of module; Carston (1996) for an overview of the issues; Samuels et al. (1999), Fodor (2000b) for massive modularity hypothesis.

often do carry commitments about the architecture of the brain, as well as main specificity, informational encapsulation, mandatoriness and cognitive impenetrability, to name just a few of their occasionally contested properties (Fodor 1983, 2000b), it is worth commenting on the version of modularity introduced in Chomsky's work, which Chomsky explicitly contrasts (2000, p. 118) with modularity restricted to input and output systems as discussed by Fodor or Marr (1982).

The empiricist tradition assumes the mind to be a blank slate that is shaped entirely by experience (apart from innate combinatorial and perceptual mechanisms). Not only does this kind of empiricism fail to account for how humans are able to bring to bear vast amounts of (largely shared) knowledge on the task of processing and categorising the world around them from the moment they are born, it also attributes to the brain less complexity than is readily accepted for far simpler parts of the rest of the body:

Empiricism insists that the brain is a *tabula rasa*, empty, unstructured, uniform at least as far as cognitive structure is concerned. I don't see any reason to believe that; I don't see any reason to believe that the little finger is a more complex organ than those parts of the human brain involved in the higher mental faculties; on the contrary, it is not unlikely that these are among the most complex structures in the universe.

(Chomsky 1979, p. 81)

Just as the rest of the body is composed of different organs, modules for Chomsky could be thought of as mental organs, the nature of which is to be discovered empirically. The language faculty is characterised by principles which may well be specific to it. Another system that has been studied successfully in this way is vision. The visual system, like the language faculty, has principles and a critical period in development. This is not to say, however, that the language faculty and the visual faculty are the same kind of system or that the principles in operation in each case are of the same sort (see Chomsky 1988, p. 161).

potential modules should be investigated in the same way, i.e. scientifically, may well not turn out to operate in exactly the same way, or have the same characteristics:

To identify these cognitive systems and to discover their properties and modes of interaction we must be willing to entertain fairly far-reaching idealization and to attribute internal structure, sometimes in the form of rules and representations, to the human mind, including substantial innate structure, which might take various forms: principles, rules, systems of representation, schemata, modes of functioning and integration, and so on.

(Chomsky 1980, p. 181)

This is where the notion of module understood architecturally comes apart from that of module understood as knowledge. An immediate question is what other modules there are beyond those for language and vision. Arguably, aspects of humans' commonsense knowledge present themselves as possible candidates. Bearing in mind that while the *methodology* that is fruitful in the case of language can be extended usefully to the study of other domains, it is not expected that other domains will succumb to exactly the same type of *analysis*, let us now consider how to investigate common sense.

2.3 LEARNING THEORY FOR ORGANISM O IN DOMAIN D: LT (O, D)

One way of thinking about the task for ethnoscience is to examine the general programme of research outlined at the beginning of *Reflections on Language* (1975), where Chomsky considers how the approach argued for and used by linguists for the study of language could have a wider application and contribute to the study of the mind more generally:

It is not unreasonable to suppose that the study of this particular human achievement, the ability to speak and understand a human language, may serve as a

uggestive model for inquiry into other domains of human competence and action
at are not quite so amenable to direct investigation.

(Chomsky 1975, p. 5)

Exactly how the study of language is to serve as a model for investigating other areas of human cognitive competence can be illustrated by introducing the notion of an $LT(O, D)$, that is, a ‘learning theory’ for an organism O in a domain D .

It may seem strange to see a reference to *learning* in the context of the study of language, when the evidence presented earlier in section 2.1 casts doubt on whether ‘learning’ is even an appropriate word to use for explaining the development of linguistic abilities, given how highly constrained this development is, and suggests that perhaps ‘acquisition’ would be more accurate. But the choice of a term like ‘learning theory’ is in part explained by the fact that the discussion in *Reflections on Language* comes into focus against a backdrop of debate between empiricists and rationalists about how to understand the mind and explain behaviour. The broad empiricist view is that the attribution of a *general learning theory* would be sufficient to explain how humans come to exhibit the behaviours they do. This expectation is questioned by a more rationalist view and replaced by the suggestion that humans develop varied and complex cognitive structures, and that any ‘learning theory’ for some domain and some organism is a function that relates the organism to its understanding of that domain via a system of (innate) knowledge that may be specific both to the domain and to the organism (see Chomsky 1975, chapter 1).

In general, instead of attempting to characterise some mechanism relating experience to behaviour *directly*, the task for a “science of behaviour” (ibid., p. 17) is to first determine how experience is mapped onto a cognitive state; only then can it move on to consider how the cognitive state attained leads to behaviour, a much more difficult question. In studying the initial question of how experience maps onto cognitive states, the possibility emerges that such mapping is facilitated in particular ways for some of the areas humans construct an understanding of, and not others.

Here humans are so facilitated in their interpretation of the events around them that they may have hit on a domain where a ‘learning theory’ is applied; examples are aspects of so-called commonsense understanding, which not only equips humans with expectations about their world but has them to a large extent agreeing about what those expectations are. We are talking here about the properties shared by humans as a species, abstracting away from individual differences.

There is more than one reason for returning to the LT (O, D) formulation here. First, it is interesting to explore an area of Chomsky’s work that is not discussed much; considering a less familiar statement of familiar ideas may prove insightful. Also it is in discussion of LTs that the areas I am interested in focussing on have been mentioned most by Chomsky. Second, the idea of a learning theory (LT) seems similar in a meaningful way to the notion of a Chomskyan module, but by talking about LTs rather than modules we avoid the usual arguments over whether or not something counts as a module and if so according to whose definition.¹³ Third, the formulation provides a framework in which to think about different aspects of the abilities we want to investigate, as I hope will become clear after I have illustrated how it works in more detail. Let us start with language, where we already know what to expect. Then we will be able to see what else this way of looking at things buys us when thinking about other domains that are less well understood.

2.3.1 *LT (H, L)*

Having fixed the organism O as human, H, a scientist S might want to establish the domains for which H is in possession of a learning theory which is specific to the domain and interestingly similar across individuals. In the case of language for

¹³ In fact, Chomsky has recently returned to this terminology: “I assume that mental capacities are ‘modular’ in the sense of Chomsky (1975), with ‘learning theories’ LT (O, D) that may vary for organism O and cognitive domain D. The resulting modules might then have input/output properties of the kind analyzed in Fodor (1983), while belonging to a ‘central’ system more structured than Fodor assumes” (Chomsky 2000a, p. 140).

the aim is to determine LT (H, L), the learning theory for humans in the
of language, which explains how they process linguistic data to arrive at the
representations they do. It is worth noting before proceeding that it is not only
'learning' that is being used loosely. 'Theory' is also being used in a way that is
slightly different from how it is sometimes strictly understood. In the philosophy of
science for example, a theory is expected to include some explicit body of axioms.
Here, a theory, the LT (O, D),

can be regarded as a system of principles, a mechanism, a function, which has a
certain 'input' and a certain 'output' (its domain and range, respectively). The 'input'
to the system LT (O, D) will be an analysis of data in D by O; the 'output' (which is,
of course, internally represented, not overt and exhibited) will be a cognitive
structure of some sort. This cognitive structure is one element of the cognitive state
attained by O.

(Chomsky 1975, p. 14)

In this sense the theory, the LT, is not, or not necessarily, a set of generalisations or
whatever is standardly associated with a theory (although, to the extent that an
explicit theory of an LT could be constructed by a linguist, that might look more like
a standard 'theory').

It is helpful to relate the ideas of LT (H, L) and its input and output to the
notions of UG and I-language introduced earlier. So, "UG is a significant
component of LT (H, L)" (ibid., p. 29). But LT (H, L) is not just the same as UG –
presumably, the LT also contains information about how to make use of linguistic
data to get from UG, the principles common to all natural languages, to particular I-
languages. I-language corresponds to the outputs of LT (H, L). The input is the
triggering experience, subjected to preliminary analysis for it to be recognised as
pertaining to the domain (in other words, the input is "the experience of O in D",
ibid. p. 15).

ow we have more ways of thinking about and referring to the different called upon in acquisition of knowledge in some domain. In what follows, I intend to call upon some combination of terms, using whichever best illustrates the issue at hand once we extend these ideas and apply them to other domains beyond language.

2.3.2 LTs for humans in other domains

Here is an initial indication of other areas where humans have skills that could be associated with a mental structure beyond general purpose reasoning:

Alongside of the language faculty and interacting with it in the most intimate way is the faculty of mind that constructs what we might call ‘commonsense understanding’, a system of beliefs, expectations, and knowledge concerning the nature and behaviour of objects, their place in a system of ‘natural kinds’, the organization of these categories, and the properties that determine the categorization of objects and the analysis of events.

(Chomsky 1975, p. 35)

In other passages, some of which have already been presented, Chomsky suggests more specific areas, like expectations about the movement of physical objects, face-recognition and social awareness. On the assumption that these areas do indeed form part of the cognitive capacity of H, that there is an interesting LT for them, i.e. one that amounts to more than trial and error or induction, several questions arise: about the initial state of the LT in each domain; about what triggering experience is needed for the competence to grow normally; and about the possible variation in final states across individuals. To draw the parallel with language again, we are interested in the equivalent to UG of whatever domain is at issue, and the final state corresponding to I-language that humans end up with, for each domain that has an LT associated with it. And we are interested in, *mutatis mutandis*, how the ‘UG-equivalent’ gets initialised and yields the ‘I-language-equivalent’ for each domain. In

are interested in the *I-structures* that arise for different domains, and the mechanisms that mediate their formation.

As the LT embodies the function that takes the child from the initial state for some domain to the corresponding I-structure, the LT will include as a sub-part the starting point for the domain. So LT (O, D) is a useful notion for thinking about both the initial state and what kind of information is required to build the output. For example, the situation for each domain could be different from how language is thought to work. Just to mention some possibilities, the initial state might just develop, or mature, into the final state without the need for any specific triggering. Or the initial state might require triggering but the evidence on the basis of which triggering takes place could be uniform, in which case there might not be even the superficial variation between final states found for different natural languages. Or again, perhaps an initial state not exposed to its usual triggers could be applied to process some other sort of information to yield a very different sort of output.¹⁴ Perhaps in some environments the relevant triggers are just not present, or not present at the right time, and the potential of the LT for some domain is not realised. Of course, these kinds of questions are entirely empirical and at this stage all of this is just speculation about alternative scenarios.¹⁵

2.4 INTRADISCIPLINARY TENSIONS

Once we have determined a domain for which humans are endowed with an LT, and we have some idea about the nature of the competence that develops for dealing with the domain thanks to that LT, there may be further issues about how the I-structure for the domain is affected by performance issues and how it interacts with other competences. What an individual gets from an LT in some domain may well not

¹⁴ Sperber (1994a) considers a possible contrast between ‘proper’ and ‘actual’ domains of modules.

¹⁵ I consider some possibilities in more detail for the area of psychology and social understanding in 5.1 below.

ll that she is able to bring to bear on the domain subsequently. Conversely, anything that is employed in understanding some domain should be ascribed automatically to the I-theory emerging from the LT (see the beginning of chapter 4). In the course of the investigation it might become clear that something initially taken to be a (cognitive) domain actually has been determined too narrowly or too broadly.

Crucially, individuating a domain, and an associated LT and respective output, is not just a matter of identifying some area of ability that humans have and positing a structure to account for that ability *understood exactly as it is described pre-theoretically*. For example, linguistic ability could be described as ‘the ability to speak and understand a language’, but, as we have seen, not all facts related to language so construed are to be explained via the language faculty directly and exclusively. In fact, how much of ‘linguistic behaviour’ will be explained by reference to I-language is an open question. Different phenomena may be re-classed as properly linguistic or otherwise according to evolving theories. Indeed, what might initially be thought of as goals for linguistic theory could turn out to be goals a theory of I-language neither can nor wants to realize. Furthermore, different accounts of language may be seen as more or less successful depending on which language-related facts they accommodate. In other words, tensions can arise within theories about how to account for data; and between theories about which data need to be accounted for. I take these two points in turn in 2.4.1 and 2.4.2, drawing together implications for common sense in 2.4.3.

2.4.1 *Interface issues*

To illustrate the changing nature of scientific linguistic theories as phenomena become better understood I will describe a picture of the state of research in linguistics and contrast claims made today with some previous assumptions.

I-language is currently conceived of as a set of principles which determine the set of well-formed pairs of structural descriptions: LF and PF. That LF and PF (‘logical form’ and ‘phonetic form’ respectively) are referred to indirectly in this way helps to remind us that the language system does not, on its own, map sounds onto

; so for instance, LF is not a representation of ‘what the uttered sentence
 LF and PF are levels of linguistic representation which interface with
 performance systems. Articulatory-phonetic (A-P) systems interface with PF
 structures. Conceptual-intentional (C-I) systems interface with LF structures.¹⁶
 Much work on the interface between I-language and A-P and C-I systems focusses
 on how a cognitive system so-configured delivers ‘speaking and understanding’.¹⁷
 For instance, semantic theories have been concerned with building accounts of how
 the conceptual system interfaces with LFs in such a way as to accommodate semantic
 constraints (e.g. compositionality of meaning).¹⁸

While integrated into a system the purpose of which is speaking and
 understanding, I-language and interface theories (such as semantic theories for
 interpreting LFs) are viewed as independent of one another by Chomskyan
 linguistics. One way to think about this *autonomy* is by analogy to a computer
 program embedded in a system which employs a number of other programs in
 realising its function. The same embedded program could be running an air-traffic
 control system or the inventory system of a supermarket, depending on how the
 program is interfaced. Similarly, that I-language is embedded in a system for speaking
 and understanding does not necessarily stem from fundamental properties of either I-
 language or communication.¹⁹ On the one hand, I-language could conceivably be put
 to other purposes; on the other, if the evolution of human cognitive capacities had

¹⁶ See Chomsky (1995, p. 2).

¹⁷ Although production is less well understood than comprehension, a predominant assumption is
 that it works like comprehension in reverse (e.g. Levelt 1989, 1999; Levelt et al. 1991).

¹⁸ Larson & Segal (1995) provide an example of a semantic framework compatible with Chomskyan
 linguistic theory.

¹⁹ In fact, one strand in the latest, Minimalist, programme of Chomskyan linguistics asks whether the
 design of I-language is optimal to its purpose, that is, whether it is a perfect solution to minimal
 design specifications, something Chomsky would regard as surprising, “hence interesting if true”
 (Chomsky, 2000a, p. 93; see also 1995, p. 5).

ided the appearance of UG, other ‘programs’ that could be used for ication might have developed.²⁰

The supposition of autonomy is manifest in current theorising both at the level of language-general UG principles and at the level of principles for specific I-languages. We can illustrate the former by considering *N* and *V* features, and the latter by considering I-language specific feature systems.

At the level of UG, the focus is on properties peculiar to language but present across languages; these include recursivity patterns and long-distance dependencies. Current theories of the principles which result in these properties are cast in terms of relations among sub-trees (phrasal constituents) which carry one of a small number of abstract features on their main terminal node. These include at least *N* and *V* features. As far as UG/I-language is concerned, these patterns could be exploited for any purpose. As it happens, in the lexicon *N* features are attributed to nominal expressions and *V* features are attributed to relational/verbal expressions. That is, the feature contrast is exploited by the system overall so that the LF structures can be interpreted as (very roughly) predicate-argument structures.

I-languages contain systems of features which can be exploited for a variety of purposes. To illustrate this idea, we can consider how I-languages might employ feature systems to regulate how surface constituents manifested in phrases are correlated with the semantic roles of the main verb. Relational verbs sub-categorize for noun phrases (and prepositional phrases) corresponding to the number of arguments they take. The lexicon must specify how to determine, on the basis of the surface manifestation of a sentence, which noun phrase corresponds to which participant in the type of eventuality the verb describes. For instance, imagine that a

²⁰ In particular, connectionist models suggest that a creature without UG but equipped with other human cognitive capacities would be able to learn certain kinds of languages. The languages in question are far simpler than actual human languages, lacking properties like recursivity, but could nevertheless serve communicative purposes.

from 'abc' is used to talk about the act of *hitting* in some language. Given a ordered string of that language 'abc NP NP', we need to be able to tell, either from surface order or morphological marking, which NP is to be interpreted as the hitter, and which as the hittee. How this works exactly depends on the version of the theory you adopt, but one basic idea (simplifying greatly for illustration) is that we suppose that UG allows for noun phrases to be marked with features $C1, C2, \dots, Cn$ from an I-language specific set. Interface systems exploit such features in coordinating which NP is for which role. Subcategorization information contained in a verb's lexical entry relates each of its semantic roles to one of its subcategorized NPs marked with an abstract feature. The C feature will either give rise to a morpho-phonological realization or the occupation of a designated place in the surface structure, and so a language user can infer which role an NP is satisfying. The lexical entry for 'abc' is suggested in (1), where the subscripts $C1$ and $C2$ signify which abstract C-feature the projections of N carry:

- (1) [s abc [NP_{C1} NP_{C2}]]
 NP_{C1} is the hitter
 NP_{C2} is the hittee

It should be clear from this illustration that, from an I-language internal perspective, the C-feature system could be being put to any purpose by the interface systems. The features themselves have no intrinsic significance. The same goes for other feature systems made possible within UG. However, from the interface perspective, the conceptual system needs to be able to 'see' these features and assign to them a significance. Correspondingly, in 'learning a language' individuals need to discover which particular configurations of feature systems are to be initialised in their I-language and the purpose local conspecifics put them to.

We are now in a position to see quite how sparse and abstract is UG/I-language according to current Chomskyan linguistics. Clearly, this kind of linguistics does not

for all pre-theoretically ‘linguistic’ generalizations. Indeed, the notions are far not only from pre-theoretic ideas about language, but also from the views of non-Chomskyan linguists. Even linguists who consider their approach to be largely Chomskyan may find themselves at odds with theoretical syntax at times. In particular, evolution in theories of I-language has resulted in situations where certain ‘linguistic’ phenomena are no longer considered part of what I-language accounts for.²¹ To illustrate, we can consider the fate of many ‘semantic’ features.

Many semi-productive lexical generalizations appear to have a semantic basis. To account for such generalizations, theorists have often presumed that I-language structures can contain ‘purely semantic’ features, i.e. features for which there is no syntactic evidence. The deployment of purely semantic features in I-language description is usually motivated by observations embodied in questions like *Why can you only ‘diagonalise’ a matrix?*²² Or *Why can’t you ‘sweep a room of dust’?* From the demise of Generative Semantics (see Newmeyer 1996), such questions have been viewed with suspicion by syntacticians as perhaps non-issues, while continuing to both fascinate and frustrate curious linguists. The idea that these facts should be captured by linguistic theory persists, it would seem, in part because of their quasi-systematicity. It is hard for linguists to believe that these phenomena are not part of what an account of language has to explain. Let us consider some examples before speculating about the source of such phenomena.

There are in fact a large number of potentially intriguing generalizations which linguists have difficulty in discounting as properly linguistic. Take for instance discussions of *verbs of removal* (Levin & Rappaport 1991), where we have classes of sets of cognate, homophonic lexical items with different distributions of subcategorization properties which apparently correspond to subtle differences in meaning. Examples of each class are given in (4):

²¹ Indeed, it is conceivable that, as the overall theory progresses, once-discounted phenomena are re-integrated among the explananda of the theory of I-language.

²² See McCawley (1976) for similar examples of what have been claimed to be purely ‘linguistic’ facts.

- a. John removed the dirt from the table.
- b. John wiped the dirt from the table.
- c. John cleared the dirt from the table.

Here, all three verbs share an [NP PP] subcategorization, and all imply that the dirt is removed from the table at the end of the action. Now consider (5):

- (5) a. John removed the table.
- b. John wiped the table.
- c. John cleared the table.

The ‘wipe’-class and the ‘clear’-class of homophone sets can participate in a further subcategorization alternative ([NP]), but (5a) cannot be used to mean that John removed something from the table (on analogy with 5b and 5c). There are further discriminations. Members of the ‘wipe’-class have a *resultative* homophone in their set which the ‘clear’-class do not (in 6), while members of the ‘clear’-class, unlike members of the ‘wipe’-class, have an *NP of NP* alternate (in 7):

- (6) a. John wiped the table clean.
- b. * John cleared the table clean.
- (7) a. * John wiped the table of crumbs.
- b. John cleared the table of crumbs.

Levin & Rappaport argue that this distribution of subcategorization frames can be predicted from ‘semantic features’ of the verbs in question. The ‘wipe’-class encode the *manner* of the described activity (‘rub’, ‘scrub’, ‘scour’, ‘brush’, ‘sweep’), while the ‘clear’ class encode aspectual information about the typical *end-state* (‘clean’, ‘empty’). But this generalization seems problematic. Consider verbs such as ‘to drain’ and ‘to strain’. Should we expect them to pattern with the manner-class (removal being with

g arrangement or a strainer)? Unfortunately, ‘manner’ is not sufficiently
 1ed by Levin & Rappaport for us to be able to tell. It so happens that ‘drain’
 and ‘strain’ seem to encode endstate information (if you have drained a glass, it is
 empty; if you have strained the tea, it contains no tea leaves). That these verbs
 encode this information probably has something to do with the fact that ‘drain’ and
 ‘strain’ pattern with the ‘clear’-class rather than the ‘wipe’-class. There is an [NP of
 NP] alternate of these forms (8c, 9c) but apparently no resultative alternate (8d, 9d).

- (8) a. John drained the water from the sink.
- b. John drained the sink.
- c. John drained the sink of water.
- d. ? John drained the sink empty.

- (9) a. John strained the tea leaves from the tea.
- b. John strained the tea.
- c. John strained the tea of tea leaves
- d. ? John strained the tea clear/leafless

The problem for Levin & Rappaport is that it is not always possible to deduce
 whether a verb of removal will have entailments about the endstate of the activity just
 from whether it is de-adjectival (like ‘cleaned’ and ‘cleared’). Nor can you do so from
 the activity involved in the type of eventuality the verb makes reference to.
 Correspondingly, we cannot tell necessarily which subcategorisation frames a verb of
 removal will take just from these indicators. Levin & Rappaport cannot tell us what
 it is about the activities of sweeping and straining such that a verb of removal
 denoting the former activity does not have the endstate entailments of the latter.

Levin & Pinker (1991) contains many more examples of attempted analyses of
 these semi-productive phenomena in terms of semantic features. Generally, the
 linguists engaged in this kind of work think of the semantic features they appeal to as
 features of I-language representations. This conflicts with the view on which LFs
 contain only features which are active in syntactic description (see Adger

ing). According to current thinking, the subcategorization properties of individual verb form are independent of subcategorization properties of semantically related or homophonous forms.

The fact that there are elements of systematicity in this area is more related to how neologisms are produced, and correspondingly, how words are learned. Suppose that you understand what *sweeping* involves, and you encounter an intransitive verb, 'pqr', which is used to describe this activity. Suppose you also already know a set of other verbs denoting similar actions which have a 'removal' subcategorization frame like that of the verbs in (4) above. It is likely that you would then be inclined to produce a (what is to you) *new* verb-of-removal with the same phonological form as the intransitive 'pqr', but with an [NP PP] frame with the NP argument linked to what is swept and the PP argument linked to where it is swept from (as in 'John pqr-ed the dust from the room'). You would be far less inclined to posit alternative lexical items with for instance an [NP of NP] subcategorization frame where the argument roles are the other way around (the first NP linked to the location of the sweeping and the second NP linked to what is swept, as in 'John pqr-ed the room of dust').

Language users are no doubt aware, in some sense, of these patterns and the inclinations they give rise to. Perhaps it is this 'meta-linguistic' knowledge that gives rise to prescriptive intuitions such as 'You can't say, "He hoovered the room of dust"'. If this suggestion is plausible, such intuitions are born of sensitivity to emergent conventions of language rather than I-language properties. I return to these issues in 5.4 below.

Because research in linguistics has made such progress over the last fifty years, it is a discipline where we are already in a position to take a historical perspective and witness the effect of theoretical shifts on how the object of study is conceived. We should bear in mind the repercussions of these developments as we turn to the investigation of common sense. Another area where the experience of linguists may

: to ethnoscientists as scientific understanding of common sense progresses, fronting pre-determined ideas about what should be accommodated by their theories. In the next section, I consider the kind of requirements theories of language have been expected to meet, and why.²³

2.4.2 *Clashing desiderata*

We have seen how we can recognise the far-reaching involvement of language in our lives without expecting a single theory to account for, say, the social function, cognitive and interpersonal significance, superficial properties and rhetorical power of language, to name but a few disparate issues. As a discipline, philosophy also has many language-related interests. Moravcsik (1975) gives various examples of why philosophy in particular has cared about language:

Some philosophers aim to draw conclusions about the structure of reality based on the alleged structure of language, others attempt the same as applied to the structure of the mind [...] others still view the study of language as helping towards self-knowledge, while there are always those who find in the key concepts of a science of language a fascinating field for general conceptual analysis.

(Moravcsik 1975, p. 13)

He goes on to note that even within the confines of philosophy more than one theory may be needed to address these interests: “clarification is achieved when we recognise that there are several distinct philosophical concerns that involve attention to language; nothing is gained by trying to explain all of these with an allegedly unified account” (ibid, p. 13). Furthermore, while linguists recognise different areas of interest in phonology, syntax, semantics and pragmatics, philosophy is predominantly preoccupied with questions about ‘meaning’. Three main reasons might at least explain, though not justify, this imbalance, according to Moravcsik.

²³ In effect, this is an illustration of the methodological dualism discussed in chapter 1.

irst, issues about meaning are historically related to issues in metaphysics.

between realists and nominalists about the existence of abstract entities can be recast in terms of how meanings are construed by different semantic theories. The theoretical constructs of syntax and phonology are not taken to carry such ontological implications. For, second, it is assumed that syntax and phonology will not posit entities which are potentially as problematic as meanings may turn out to be: “many philosophers feel that while it is relatively clear what sounds are and what segments of language are, it is far from clear what meanings are” (ibid., p. 16).²⁴ A final factor in philosophical semantic-centricity is a widely, if quietly, held belief that once meaning has been understood other language related problems will fade away.

These views may rarely be explicitly stated, yet there was a time when something like them featured widely in many philosophers’ thinking “if not in their programs at least in their practice” (ibid., p. 14). Today this is probably a lot less the case; thirty years on it is to be hoped that even if theorists from different perspectives are still reluctant to change their starting points, they are at least aware of the aims and objectives of other approaches. To what extent a lesson has been learned for the study of other aspects of the mind remains variable across fields and individuals.²⁵

2.4.3 *Looking forward*

In a weak sense, then, there is a tension between the study of language and the study of common sense simply because of all the ground that the former has already covered; the latter has to catch up.

First, I have considered a recent picture of the language faculty and linguistic research. Second, I have discussed how I-language can be related to ‘speaking and

²⁴ In the light of the discussion of, for instance, syntactic features in 2.4.1 above, this kind of assumption further unmasks the potential lack of appreciation for the sophistication of the knowledge structures underlying our linguistic competence. See also 5.2.2.

²⁵ I will argue that there is a comparably deeply engrained orthodoxy in some discussions of commonsense psychology; it can take considerable effort to shake up predominant assumptions in the face of entrenched arguments unquestioningly presented.

iding', noting that particular data might be considered specifically syntactic, : of the application of interface systems, or independent of the language faculty, depending on the state of the theory and how boundaries are drawn. Third, I have also acknowledged the need to be aware of other constraints on accounts of language and how these arise. All these points will be relevant as we turn to the study of commonsense psychology, in chapter 4; I take them in reverse order here.

Theorists of commonsense psychology (philosophers and psychologists both) are well aware of the third point, i.e. that there are a variety of questions at issue.²⁶ However, not much is done about the consequences for debate of this situation beyond recognising what the different questions are, at best. People tend just to discuss the ones they are interested in without justifying the subset they choose, or without acknowledging that the basis on which they choose some subset might conflict with assumptions motivating the choice of another. With regard to the second point, people realise that the boundaries of what they can account for, and how, may shift; but they often lack an explicit conceptual and methodological framework to spell out the consequences of their choices.²⁷ It is as if syntacticians were to try and do syntax not only without anyone having taken on philosophers on their behalf, but also secretly be seeking their approval. The problem is not so much a failure to engage, as a failure to disengage constructively. Research in this area lags behind the study of language (certainly there is little as sophisticated as the programme so briefly described under the first point above), in part because it is hampered by these problems. By addressing the imbalance it becomes easier to recognize what kind of contribution is being made by particular proposals and thus easier both to defend and evaluate them on their own terms.

²⁶ Stone & Davies (1996) distinguish nine, including a concept mastery question, a metaphysical question, and descriptive and explanatory questions about adult practice and development.

²⁷ For instance, within developmental psychology the significance of specific experiments like false-belief tasks may be reassessed, with consequences across the philosophical divide which are not always pertinent to the reassessment. These points are addressed in detail in chapter 4 below.

3

ON FOLK SCIENCE, EARLY SCIENCE AND THE SCIENCE-FORMING FACULTY

3.1	ASPECTS OF ETHNOSCIENCE.....	72
3.2	LEARNING THEORIES (LTs) AND NATURAL SCIENCE.....	75
3.3	A FACILITY FOR SCIENCE; A FACULTY FOR SCIENTIFIC REASONING.....	78
	3.3.1 <i>Abduction in SFF and LTs</i>	82
	3.3.2 <i>Abduction and evolution</i>	85
3.4	EARLY SCIENCE.....	87
3.5	THE SUCCESSFUL NATURAL SCIENCES.....	89
3.6	OVER-EXTENSIONS AND LIMITATIONS.....	91
3.7	ETHNOSCIENCE IN THE CONTEXT OF WIDER SCIENTIFIC INQUIRY.....	94
3.8	TERMINOLOGICAL NOTE.....	99

3

ON FOLK SCIENCE, EARLY SCIENCE AND THE SCIENCE-FORMING FACULTY

There are now three main areas to address that are separate but interconnected: first, the relation between various aspects of the conceptual resources of folk sciences and the activity of scientific inquiry; second, the range of naturalistic ethnoscience; third, the place of ethnoscience in the context of wider scientific inquiry. These strands relate to a main theme of the chapter, the scope and limits of naturalistic inquiry.

3.1 ASPECTS OF ETHNOSCIENCE

Let us consider a statement of the focus of ethnoscience and the issues that face it:

One branch of naturalistic inquiry studies common-sense understanding. Here we are concerned with how people interpret object constancy, the nature and causes of motion, thought and action, and so on ('folk science', in one of the senses of the term). Perhaps the right way to describe this is in terms of beliefs about the constituents of the world (call them 'entities') and their organization, interaction, and origins. Assume so. It is an open question whether, and if so how, the conceptual resources of folk science relate to those involved in the reflective and self-conscious inquiry found in every known culture ('early science'), and to the particular enterprise we call 'natural science'. For convenience, let's refer to the study of all such matters as 'ethnoscience.'

(Chomsky 2000, p. 135)

Clearly, there is much to absorb from this passage. In what follows, I aim to explore some of the ideas touched on and how they relate to existing positions. Chomsky goes on to note that : “It is also an open question how the conceptual resources that enter into these cognitive systems relate to the semantic (including lexical) resources of the language faculty.” I return to this remark about linguistic resources and the concepts of commonsense understanding in chapter 4, although it will be useful to bear it in mind as we proceed to address the other main questions being raised here: (i) the nature of commonsense understanding, (ii) its relation to early science, and (iii) its relation to natural science. While we may refer to the study of all this as ethnoscience “for convenience”, it is important to recognise the three elements.

The first of these is the most familiar aspect of what we have been calling ethnoscience. There is both a positive and a negative reason for now including the other two. The positive reason is that if the resources of folk science do inform early science and affect natural science then it is part of understanding the reach of folk science to establish what such relations consist in. The negative reason is that it is necessary to underline the fact that neither the existence nor the essence of any such relations can be assumed. Now we can take the points of the quote in turn:

(i) One branch of naturalistic inquiry studies common-sense understanding. Here we are concerned with how people interpret object constancy, the nature and causes of motion, thought and action, and so on ('folk science', in one of the senses of the term). The sense of ‘folk science’ that is relevant to us here is re-iterated in the quote: folk science is what is studied by the branch of naturalistic inquiry concerned with commonsense understanding. We might also note how this sense is hardly evoked by the terminology widely adopted. The natural language word *folk* carries many connotations: for example, it is associated with traditions, cultural practices, hearsay. Labelling some piece of knowledge as ‘folk knowledge’ would normally tend to carry implications about its origin and application that are related to these kind of connotations. If we are trying

to understand the nature of ‘folk knowledge’ in the sense of commonsense understanding we have been discussing, then where such knowledge comes from, how it is passed on, how widely it is entertained, and how explicitly, and how reliable such knowledge can be expected to be, are all questions that need to be addressed. Without wanting to prejudge the issue, the impressions created by the use of *folke* point us in the opposite direction from that in which we are heading; remember the aim is to discover the shared traits of our cognitive capacity, not folkloristic bits and pieces of information picked up haphazardly and optionally. In 3.8 below I consider other senses of the term ‘folk science’ and briefly contrast what we are interested in with homonymous activities. For the moment let us acknowledge that as the term is now widely used it needs to be clarified rather than just abandoned.

One reason the term may have gained currency originally is to distinguish scientific theories from folk theories – scientific physics from folk physics, for example. But, as we saw in chapter 1, this is a contrast that needs to be emphasized if the object of inquiry is that which the scientific theories and the folk theories are directed *at*. If folk theories themselves are what are under investigation, then how folk theories compare to scientific theories with regard to what they are theories *of* – in other words, what the status and predictive success are of folk science as opposed to natural science with regard to some phenomena – is of secondary importance, if any, to establishing what the folk theories actually amount to. Furthermore, once folk theories *are* themselves what are under investigation, then we still need to be able to distinguish between folk and scientific accounts of them, but talking about, e.g., scientific folk physics and folk folk physics is cumbersome, at least, and the two risk being collapsed one onto another.

This further potential for confusion supplies us with an added incentive to introduce more specific terminology where possible in future, and at least to examine any occurrence of ‘folk’ with care.

(ii) *It is an open question whether, and if so how, the conceptual resources of folk science relate to those involved in the reflective and self-conscious inquiry found in every known culture ("early science").* So, what might be the relation between folk science in our favoured sense, and early science? The notion of early science will be considered in more detail in 3.4. As described briefly here, early science sounds like the sort of thing that might fall under a wider construal of 'folk', for example particular traditions or philosophical analysis. Do LTs contribute to early science? Perhaps early science, with its culture specific notions and reflective ideas, could be a side-effect of the presence of LTs for some domain; early science might yield conscious representations that are constrained at a deeper level (maybe by the relevant I-theory).

(iii) *It is an open question whether, and if so how, the conceptual resources of folk science relate to ... the particular enterprise we call 'natural science'.* This is the question to which I now turn. Scientific knowledge may be a different kind of knowledge from that provided by folk science; but perhaps it exists in virtue of the same principles.

3.2 LEARNING THEORIES (LTs) AND NATURAL SCIENCE

To gain further understanding of the general nature of the human mind, we should ask in what domains humans seem to develop complex intellectual structures in a more or less uniform way on the basis of restricted data. Wherever this is the case, we can reasonably suppose that a highly structured genetic program is responsible for the achievement, and we can thus hope to learn something significant about human nature by studying the systems attained.

(Chomsky 1980, p. 248)

This statement sums up the program we are pursuing; put simply, in the terminology introduced in the previous chapter, we are looking for learning theories (LTs). One

area where we take them to be in operation is language. Having established that, we are now on the lookout for LTs in areas of commonsense understanding.²⁸

However, it is not only through the possession of LTs that humans have at their disposal ways of filtering and accommodating information about the world. Humans sometimes address problems for which they may have an LT, but attempt to do so from outside the perspective provided by (the output of) that LT; and they also attempt to address problems for which they have no specific LT. Both of these scenarios are instantiated in the domain of physics. For an example of the first case, it appears that humans do possess some kind of LT for physics (giving rise to commonsense understanding of the phenomena physics explains), but when attempting a scientific account of physical phenomena the intuitions of commonsense physics may be dropped, and the view provided by the LT eventually replaced by a scientific stance. For an example of the second case, research in physics goes beyond the areas of physics for which the relevant LT equips us, so humans engaged in such research will be addressing questions on which specific LTs are silent.

In fact, it is unlikely that we need to look as far afield as the outer fringes of physics to find such examples. There are probably many questions and areas that humans are concerned with but for which they are unlikely to be equipped with an LT. (As an aside, we might note that among them are those questions with which this thesis is most concerned, specifically, the nature of LTs themselves.) Yet, I have been arguing, given the precedent set by success in the realm of the study of LTs in the case of language, aiming to characterise other possible LTs within this paradigm is not such a wild ambition. Despite the absence of dedicated LTs to tackle many

²⁸ 'LT' is used loosely here as a way of referring to both an initial state of the mind and its development into an understanding of some area, and to whatever resulting knowledge structure shapes subsequent judgements about that area. Speaking more precisely, having identified an area that might be subserved by an LT we can attempt to characterise the knowledge structure available for that area – the I-theory – and the process by which it arises thanks to the LT.

questions of scientific interest, impressive progress has been made in specific areas, and we may be hopeful about prospects for others.

Still, the knowledge that comes from explicitly constructed scientific theories (including those that might be put forward in linguistics and ethnoscience) can be distinguished from knowledge that arises naturally in humans for language and common sense. Considerations to do with universality, origin, ease and speed of acquisition should by now be familiar, and apply to language and common sense in a way they cannot be taken to hold of the development of scientific knowledge. Conversely, coming up with a decent theory generally takes individual brilliance and creativity coupled with fruitful collaborations; the process is very different from the acquisition of language and common sense not only because individuals engaged in it are rare, but also because successful science is usually the result of many different individuals working together over time.

There are some similarities, though, between knowledge of language and of common sense on the one hand, and the development of scientific knowledge on the other. Usually this observation is made by psychologists claiming that the development of commonsense knowledge follows the pattern of theory change in science, with genius-like children passing through successive scientific revolutions before miraculously alighting on the same account of their environment (see e.g. Wellman 1990; Gopnik & Wellman 1992; Gopnik & Meltzoff 1997). However, this is not a proposal I would subscribe to as it leaves unexplained how all normal children succeed in doing this despite undeniable differences in their individual capacities and in the available evidence they can bring to bear on their theorising (see Segal 1996). It also fails to accommodate facts about dissociations between abilities

in particular areas of common sense, and the capacity for the kind of general purpose reasoning that is alleged to be called upon in their development.²⁹

Perhaps, though, it might be worth pursuing the comparison in the opposite direction. Rather than comparing commonsense capacities to the scientific process, then, let us consider whether there is an interesting way of comparing the scientific process with commonsense capacities.

3.3 A FACILITY FOR SCIENCE; A FACULTY FOR SCIENTIFIC REASONING

One way of thinking about what an LT for some domain does, is that it constrains possible hypotheses about how the information it deals with should be organised. Without the contribution of LTs, it would be hard to see not only how people arrive at similar representations of the world, but how, even individually, they could reach many conclusions at all, given the amount of potentially relevant information available to them and the vast spectrum of possible ways of interpreting and arranging it.

In a way, a similar situation holds for scientific knowledge: there appear to be constraining principles on possible moves to be made in developing a scientific theory. Without such constraints it would be hard to see how theorists could hit on anything in the face of so many options. And it is striking given quite how many possibilities there are that only some of them are entertained seriously (or at all) and put forward as candidate explanations for some set of phenomena. It is also striking that different scientists often converge independently on the same kind of theories.

Now, in the light of previous considerations, it would be implausible to claim that science depends on LTs in exactly the same way that language and common sense do.

²⁹ Individuals with Williams' syndrome, for example, appear to have an intact understanding of psychological categories without being able to carry out other basic inferences; see e.g. Bellugi et al. (1994); Karmiloff-Smith et al. (1995); Segal (1996); Botterill & Carruthers (1999) for discussion.

After all, nearly everyone gets language and common sense, but not everyone ends up doing science. Of course, everyone gets language and common sense only under the right conditions – perhaps in the case of science the right conditions are all too often just not met and given the appropriate opportunities more people would succeed as scientists. But still, the conditions required for language and common sense are minimal, thanks to the complexity of what they are mapping on to, and are present in our immediate surroundings, whereas those needed for doing science may be more disparate and less available. However, even people not capable of coming up with scientific theories in the first place, either through lack of skill, or through lack of background information, time and resources, are capable of appreciating them and tend to judge them according to widely accepted standards.³⁰

So, perhaps there are constraints on the kind of explicit theories humans are able to come up with. We might be reluctant to think of these constraints as *limits*, because they are in fact what help to provide humans with the scope of understanding we actually have, just as much as being what circumscribes our range. Without even these constraints the options would be just so many and so diverse as to result merely in confusion. For the theories we generate are grossly underdetermined by the available evidence – for any proposal, many alternatives are equally compatible with the (often patchy) data that we attempt to accommodate. Yet most of these will not even occur to us, let alone be thought plausible. As we generally are able to move beyond helplessly contemplating a morass of undifferentiated unanalysable evidence in our theorising, the next question that arises is, exactly what kind of enabling constraints are we talking about? Chomsky puts the question like this:

³⁰ For example, while people may not be good at recognising immediately whether some experimental result is statistically significant without being able to conduct the necessary mathematical analysis, the idea of statistical significance and why it is important, together with the notion of control groups, is assimilated readily.

What is the ‘science-forming capacity’ that enables us to recognize certain proposed explanatory theories as intelligible and natural, while rejecting, or simply not considering a vast array of others that are no less compatible with the evidence? I do not speak here of the creative achievement, but rather of the appreciation of the achievement, a common human ability; the ability to recognize, with understanding and pleasure, that an intelligible explanatory theory has been produced. Some such science forming capacity must be an innate property of mind. That is not to say that all potential scientific knowledge is ‘preformed’ at birth. Rather, the human mind is endowed with some set of principles that can be put to work when certain questions are posed, a certain level of understanding has been achieved, and certain evidence is available, to select a narrow class of possible theories.

(Chomsky 1980, p. 250)

More recent references in Chomsky’s writing to humans’ scientific abilities introduce the notion of a *science-forming faculty*, SFF.

Among the aspects of the mind are those that enter into naturalistic inquiry; call them ‘the science-forming faculty’ ... Equipped with SFF, people confront ‘problem situations’ ... Sometimes [SFF] provides ideas about how the questions might be answered or reformulated ... ideas that can then be evaluated in ways that the SFF offers (empirical test, consistency with other parts of science, criteria of intelligibility and elegance, etc.).

(Chomsky 2000, p. 82)

We should note that elsewhere Chomsky explicitly mentions that this label serves to “dignify ignorance with a title” (ibid., p. 22). This indicates, perhaps, that we should be cautious about automatically carrying over the attributes of the language *faculty* to this notion of a science-forming *faculty* – despite the apparent similarity in terminology, it would be hasty to assume that they are the same kind of system. Indeed, at this stage it may be more helpful to think of SFF as a faculty for science in the sense of *facility* or propensity for science; as understanding of it proceeds we can move on to consider what claims to make with regard to knowledge structures.

While the “science-forming capacity” (as it is also referred to) is seen as another “particular component of the human biological endowment” (Chomsky 1988, p. 158), Chomsky contrasts it with language:

In the case of language there is a special faculty that is a central element of the human mind. It operates quickly, in a deterministic fashion, unconsciously and beyond the limits of awareness and in a manner that is common to the species, yielding a rich and complex system of knowledge, a particular language. *For problem solving and theory construction there is nothing so specific.*

(Chomsky 1988, p. 157, my emphasis)

Language and common sense are (or give rise to) cognitive structures that can be distinguished from scientific knowledge like, for example, knowledge of physics. The fourth chapter of *Reflections on Language* reiterates this difference at some length from the point of view of the unencumbered Martian scientist studying the types of cognitive structures humans appear to have. “Though [knowledge of physics] too, is derived on the basis of specific properties of mind, it does not reflect these properties in the same way as language and common sense do” (Chomsky 1975, p. 157). This is an important distinction. In the terms adopted here, language and common sense are underwritten by ‘I-theories’; by contrast, knowledge of physics is not an I-theory, however much it may depend on the science-forming faculty. If anything, it is the science-forming faculty itself that should be thought of in terms of I-theory, albeit with some reservations as highlighted above. To put it another way, what you get from the application of the language faculty is an I-language. If there is an in-built mechanism of some sort for doing science, then presumably what such a mechanism provides is not science itself but the principles for doing science and an indication of the circumstances under which it, as a mechanism, might come into play. It is, after all, a science-*forming* faculty, not a science faculty. Once the principles for doing and appreciating science, as dictated by SFF, are in place, then the ability that the faculty

gives rise to can be applied to its domain and do science, giving rise to things like knowledge of physics.

3.3.1 *Abduction in SFF and LTs*

What might these principles, according to which we operate (thanks to the endowment that is our SFF), be? And are they specific to science-forming or do they apply more generally? Let us take these questions in turn. One possibility about what might be in operation in SFF, and perhaps beyond, echoes some ideas due to Peirce. As cited by Chomsky:

[Peirce] held that innate limitations on admissible hypotheses are a precondition for successful theory construction, and that the ‘guessing instinct’ that provides hypotheses makes use of inductive procedures only for corrective action. [...] Peirce regarded inductive processes as rather marginal to the acquisition of knowledge; in his words, ‘Induction has no originality in it, but only tests a suggestion already made.’ To understand how knowledge is acquired, in the rationalist view that Peirce outlined, we must penetrate the mysteries of what he called ‘abduction’, and we must discover that which ‘gives a rule to abduction and so puts a limit upon admissible hypotheses’.

(Chomsky 1968, p. 78-79)

Abduction is a type of non-demonstrative inference. An abductive inference is an inference to the best explanation, or, if not the best explanation, something that accommodates or makes sense of a fact that is otherwise surprising.³¹ Peirce distinguished this kind of backwards reasoning from induction, another type of non-demonstrative inference where one reasons from a series of instances to a generalisation. Both of these kinds of probable inferences are contrasted with deduction, a necessary inference via which conclusions are drawn reliably from initial

³¹ Also known as ‘retroductive inference’ or ‘hypothesis’; Burks (1946, p. 301) discusses in detail “what Peirce thought to be sufficiently distinctive about abduction to warrant calling it a new kind of argument”.

hypotheses using logical rules. These three types of inference become integrated into the Peircean view of scientific method together with his notion of economics of research, whereby recognisably limited resources for scientific activity are allocated to achieve the greatest increase in knowledge for the investment made. Abduction serves to construct hypotheses, deduction serves to derive their consequences and induction relates to experimental testing of hypotheses (see Peirce 1957, p. 235-55; and Stanford Encyclopaedia of Philosophy). Abduction is thus crucial to the acquisition of knowledge, and according to Peirce can be accounted for by appeal to evolution (see section 3.3.2 below). Before commenting on this aspect of Peirce's proposal, let us turn to the question of whether the principles of SFF are specific to science or apply more widely.

At the beginning of this discussion, we noted that there surely are problems that humans address in the absence of an enabling LT. There are presumably many such problems, not all of which would naturally be thought of as falling in the domain of 'science'. If what helps us to address most problems is some aspect of SFF, then SFF is applicable beyond strictly just *scientific* reasoning, however that may be defined, and could apply also to everyday questions. If plausible, this point is quite significant and underlines why it is worth exploring a notion of SFF – it may be something that comes into play frequently, rather like a capacity for reflective thought, not just restricted to a secluded area of cognitive activity, scientific theorising.

A more substantive point about the reach of principles of SFF concerns the suggestion that some notion like abduction may be of significance not only to the development of science, where its need is widely recognised, nor just to the general problems for which humans lack an LT as we have just speculated, but also for, as Chomsky puts it, the “constructions of ‘common-sense understanding,’ that is, for our ordinary notions concerning the nature of the physical and social world, our intuitive comprehension of human actions, their ends, their reasons and their causes, etc.” (Chomsky 1979, p. 72). I am not sure what is meant here by “constructions”,

exactly. It could be understood as the view or understanding ‘constructed’ by the different elements of common sense – after all we surely do need to allow for the outputs of various commonsense I-theories to combine somehow and ‘construct’ an overview, and SFF might be involved in such a process. The skill here would be the integration of information from different sources and the evaluation of the significance to a wider picture of individual pieces of information. There must be some constraints on how humans do this given that they not only possess similar commonsense theories, but they appear to apply them similarly in extremely diverse situations to much the same effect. Perhaps the successful operation and application of our various I-theories involves some kind of abduction.³²

As an alternative, we could understand “constructions of common sense” just as meaning the different elements of commonsense understanding as ‘constructed’ by LTs. On this interpretation, the notion that a process radically different from induction is required to account for the “constructions of commonsense understanding” might mean that abductive inferences play a role in the *acquisition* of parts of commonsense knowledge; in which case, perhaps abduction features within LTs themselves. In a weak sense, this observation brings us full circle back to where we started, as it was an initial claim that common sense, and language for that matter, are acquired through LTs, and that what LTs do is constrain possible ‘theories’. In a stronger sense, however, we might ask whether a process of abduction could actually cast light on how LTs work.

Chomsky does mention non-demonstrative inference in the context of acquisition of language. For example:

[K]nowledge of a language – a grammar – can be acquired only by an organism that is ‘preset’ with a severe restriction on the form of a grammar [...] The child cannot know at birth which language he is to learn, but he must know that its grammar

³² Such an interpretation of this particular quote raises an interesting issue and one that at some point will need to be addressed. Chapter 5 considers some ideas about the integration of information between commonsense psychology and language.

must be of a predetermined form that excludes many imaginable languages. Having selected a permissible hypothesis, he can use inductive evidence for corrective action, confirming or disconfirming his choice. Once the hypothesis is sufficiently well confirmed, the child knows the language defined by this hypothesis.

(Chomsky 1968, p. 78)

And in *Aspects of the Theory of Syntax* (1965), Chomsky discusses highly constrained search spaces and evaluation procedures for grammar, which might limit the possible hypotheses to a single one (see Chomsky 1965, p. 35-45). Of course, for LTs constraints would probably be different in each domain.

Papers in Sperber, Premack & Premack (1996) deal with inference to the best explanation and backwards inferences to causes in different domains. Work by Gigerenzer and colleagues, investigating “fast and frugal heuristics” for non-demonstrative inference (see e.g. Gigerenzer & Todd 1999; Gigerenzer, Czerlinski & Martignon 2002) is suggestive in this area and has the potential to be applied widely (Gilovich, Griffin & Kahneman 2002). One aspect of some of these proposals that is not always automatically in tune with the picture developing here, however, is the kind of appeal made to evolution. Peirce also ties his thoughts to evolutionary considerations, and in the next sub-section I refer to him to highlight possible reasons for scepticism.

3.3.2 *Abduction and evolution*

According to Peirce, “man’s mind has a natural adaptation to imagining correct theories of some kinds” (Peirce 1957, p. 238). Arguably, however, a reformulation of this statement might be easier to defend.³³ Let us consider changing it to “man has a

³³ Reading on, it becomes clear that Peirce shares my concerns: “But if that be so, it must be good reasoning to say that a given hypothesis is good, as a hypothesis, because it is a natural one, or one readily embraced by the human mind. *It must concern logic in the highest degree to ascertain precisely how far and under what limitations this maxim may be held.* For of all beliefs, none is more natural than the belief

natural adaptation to imagining theories of a particular kind”. Whether such theories are *correct* (i.e. true, presumably) is then recognisably a separate issue, as is whether for any theory the imagining of it can be shown to be a natural adaptation. And looking at what makes the theories that man’s mind comes up with of a *particular* kind (if they share any features or properties for example), regardless of whether they are correct, may still tell us a lot about the kind of mind that devised them. The revised formulation captures the following points.

For areas where man is endowed with an LT we can say that man imagines the (I-) theory that the LT guides him towards – the issue of whether or not it is actually a correct theory arises only secondarily if we are interested in that which it is a theory of. Of course, in some areas, we might want to say that there is an evolutionary advantage to this theory indeed being close to correct about what it is a theory of, at least for the purposes it is used for on a daily basis (take, for example, the areas of physics and psychology covered by the LTs of common sense). However, whether the theory would need to be *exactly* correct is doubtful; an approximation would probably be sufficient for problems of everyday concern – which are presumably the ones the theory constructed by the LT evolved to tackle in the first place (indeed, the discrepancies between commonsense physics and scientific physics tend to be most salient as we progress beyond the realm of direct experience). Furthermore, in some cases, most notably language, the issue of whether the “imagined theory” – in this case I-language – is a “correct” theory does not arise at all, as there is nothing that it is a theory of, and against which the theory could be evaluated, beyond what is determined by the theory itself.³⁴

For the cases where no LT is in operation, again it is an overstatement to claim that man has a natural adaptation to imagining correct theories of some kind.

that it is natural for man to err. The logician ought to find out what the relation is between these two tendencies.” (1957, p. 238-239, my emphasis) I hope the reflections that follow count as a first step.

³⁴ This point is expanded in detail in 5.2.1. Of course, for a linguist trying to characterise the I-theory, as opposed to the individual acquiring it, there is an issue of doing so correctly.

Rather, we should say that the theories he imagines will be constrained (or guided, depending on which way you look at it) by the principles in virtue of which he operates (say, for the sake of argument, by some form of abduction plus constraints on processing). There is no guarantee that the theories available via such reasoning will turn out to be correct ones. Nor is there any reason to believe that evolutionary considerations would favour correct theories when it comes to the activity of natural science. The production of scientific theories is unlikely to be an adaptive enterprise, at least not on an individual scale. While Chomsky buys into Peirce's idea that when it comes to science there are ideas and hypotheses that are more accessible and intuitively pleasing to humans, he is not convinced that this is an evolved capacity: "while man's mind is no doubt adapted to his requirements, there is no reason to suppose that discovery of scientific theories in particular domains is among the requirements met through natural selection" (Chomsky 1975, p. 156).

3.4 EARLY SCIENCE

We have been considering how the resources for engaging in scientific inquiry reflect the resources of common sense, or the folk sciences, insofar as both kinds of understanding are constrained by our cognitive structure (though not necessarily constrained in the same way). While we are thinking about what the conceptual resources of folk science have in common with scientific activity and how the two are related (speculating that abduction may be involved in both), we might wonder about the following question: Do the resources of natural science further reflect the resources of folk science in the sense that the natural science of some domain is constrained (partially or irreversibly) by the folk science of that domain? In other words, is there a sense in which there is continuity between folk science and natural science in the areas that humans address from both perspectives?

This is a question that is important for the areas of common sense that I am particularly interested in addressing in the next chapters, and the answer to it appears

to have been taken for granted in influential statements. For example, “science is continuous with common sense, and the ways in which the scientist seeks to explain empirical phenomena are refinements of the ways in which plain men, however crudely and schematically, have attempted to understand their environment and their fellow men since the dawn of intelligence” (Sellars 1963, p. 183). I think this position needs to be qualified. I will set out some initial reasons below and return to the issue of continuity in the discussion of commonsense psychology in the following chapters.

It is interesting that the first attempts of natural science to characterise some domain tend to be approximations to the predictions of folk sciences. For example, take the widely cited fact that the knowledge that folk physics equips us with is roughly equivalent to medieval impetus theory (Stich 1996 Ch 1; McCloskey 1983; see also Hayes 1979, 1985; Smith & Casati 1994). Apparently humans can be shown to predict and explain the behaviour of physical objects through appealing to concepts like *force* and *mass*. It is claimed that humans’ movement-of-physical-object-explaining behaviour can be best accounted for by psychologists on the assumption that such concepts are the ones in operation providing humans’ intuitive understanding of the physical world. If this is true, we can make a number of comments.

For the sake of discussion, say that medieval impetus theory was developed by some studious members of a monastery. Now, it is not as if before the monks set out their theory neither they nor anyone else was able to predict and react to the movements of things around them. In a similar vein, children have folk physics but surely do not think about it in the terms introduced by the monks, if they even think about it explicitly at all, which is doubtful.³⁵ For that matter, I expect adults are not always sure about the role of concepts like *mass* and *force* if they are questioned about them; yet, on the basis of their behaviour in situations where their folk physics is

³⁵ Work by Baron-Cohen and colleagues (e.g. Baron-Cohen et al. 1999, Baron-Cohen 2000) suggests a heightened mastery of physics concepts in individuals with Asperger’s Syndrome, who perhaps do reflect on this more than normal.

called upon, they clearly rely on and are able to manipulate such concepts. In a sense, then, medieval impetus theory pre-dates, or is available independently of, the monks.

Yet presumably the medieval scientists responsible for the development of impetus theory took themselves to be saying something new about the behaviour of physical objects which up until then had not been explained. In fact, I believe, it is possible that what the monks were actually doing when they put forward their theory was providing an explicit characterisation of humans' intuitive expectations (although they would not have seen themselves as being so engaged). Thus perhaps early science of some domain tells us more about the psychology of the folk sciences than it does about the phenomena at which it is directed. Now let us consider where this would leave us with regard to the question about whether folk sciences constrain natural science.

While folk physics provides a starting point for investigation into physics, forming the basis of initial inquiry, science develops far beyond its reaches. Indeed, quantum physics cannot even be applied straightforwardly to the medium-size objects at the level at which folk physics operates (as shown by Schrödinger's cat). So while folk physics may have informed progress in science, it does not constrain it irremediably. The situation may be different for each domain we consider, but in general, following on from the arguments presented in the previous section, I expect that the main thing that constrains the success of natural science of some domain is whether correct hypotheses about it happen to be among those that the guiding principles of SFF can point us towards eventually.

3.5 THE SUCCESSFUL NATURAL SCIENCES

These considerations suggest a somewhat deflationary view of how we might characterise what the successful natural sciences are. The idea would be that sometimes we are quite near the mark simply because the kind of theories we are capable of constructing happen to fit in with how the world actually is. So the

successful natural sciences are given by the (chance) intersection between the hypotheses that our SFF directs us towards and reality; they are those where there is an overlap between the facts of the natural world and the resources of the science-forming faculty (see e.g. Chomsky 1980, p. 251, 1988, p. 157-159, 2000, p. 83; for other views of the cognitive foundations for science and scientific reasoning see contributions in Carruthers, Stich & Siegal 2002).

This point relates back to the issue of problems and mysteries introduced in the first chapter, where we set out the idea that there are probably questions humans are equipped to tackle (problems) and questions humans lack the resources to address with any degree of success (mysteries). We are now in a position to think about what explains this distinction. By studying SFF in more detail it might be possible to learn something about what questions fall in which category. “It is conceivable that we might discover the principles that underlie the construction of intelligible theories, thus arriving at a kind of ‘universal grammar’ of scientific theories. And by analyzing these principles, we might determine certain properties of the class of accessible theories” (Chomsky 1980, p. 251). Problems would fall within the scope of SFF; mysteries outside it.

Of course, the limits of SFF are the limits of scientific understanding, not the limits of any understanding at all. “The science-forming capacity is only one facet of our mental endowment. We use it where we can but are not restricted to it, fortunately.” (Chomsky 1988, p. 159) We have other ways of thinking about questions that may be beyond scientific investigation, many of them better suited than science to the topics they aim to cast light on:

Naturalistic inquiry is a particular human enterprise that seeks a special kind of understanding, attainable for humans in some few domains when problems can be simplified enough. Meanwhile, we live our lives, facing as best we can problems of radically different kinds, far too rich in character for us to hope to be able to discern explanatory principles of any depth, if these even exist.

(Chomsky 2000, p. 115)

Far from being negative or pessimistic, this position merely recognises that not every question has an attainable answer for humans, that science can cast light on some areas, and that other approaches can also be illuminating for both these and other areas that lie beyond the reaches of science.

3.6 OVER-EXTENSIONS AND LIMITATIONS

There is a further, slightly different, way science has to recognise its limits, and that is that science should not overstate its claims. The reach and uses of scientific ideas and methodologies need to be carefully stated and delineated. Scientific inquiry should acknowledge the limits of its own success. Instead, sometimes strong results lead to over-generalisation. Failure to recognise this pitfall can lead to what Dupré calls *scientific imperialism*, “the tendency to push a good scientific idea far beyond the domain in which it was originally introduced, and often far beyond the domain in which it can provide much illumination” (Dupré 2001, p. 74). Such unchecked over-generalisation goes against good science, yet, ironically, the results of it hide behind a veneer of scientific respectability inherited from the rigour of the original idea.³⁶

Clearly, if scientific imperialism is a problem, the thesis I am putting forward, about how the study of linguistics can contribute to the study of the mind more generally, needs to proceed with caution. There is a worry otherwise that it might sound as if I am pursuing an ‘imperialistic’ agenda, so it is important to be clear why this is not the case. My idea is not to carry over blindly results obtained in linguistics to other aspects of the mind, but to explore the potential for a similar approach in other domains. It is explicit in Chomsky that the separate ‘faculties’ of the mind

³⁶ In earlier publications, Dupré warns of the dangers of “scientism” (see e.g. 1993, p. 221-224). In a very different use of the term, Fodor calls “scientism” the position that holds “on the one hand, that the goals of scientific inquiry include the discovery of objective empirical truths; and, on the other hand, that science comes pretty close to achieving this goal at least from time to time.” (Fodor 1998, p. 189). In this sense, unlike the scientism of scientific imperialism, scientism is good.

tentatively suggested neither operate by the same principles (see the way vision is contrasted with language) nor necessarily determine their outputs to the same degree (see the way language is contrasted with science-forming). Even the comparison between faculties of mind and organs in the body sustains the idea that very different kinds of systems are envisaged. Faculties of mind need be no more similar than are heart, lungs and liver, they just, at least initially, are to be studied in the same manner. In short, it is not the same idea that is being repeated, but variations of it according to evidence for different domains.

Part of the reason for Dupré's concern about over-generalisation stems from his reservations about the unity of science and thus about whether the same approaches can be applied to all different kinds of questions.³⁷ A different worry, more central to this discussion, is that the effects of scientific imperialism can be particularly pernicious when the original proposals are *misapplied*.

It is important to recognise that unless, together with enthusiasm for the approach, the methodological aspects of the Chomskyan paradigm are respected, we run the risk of an unwarranted and indiscriminate proliferation of postulated LTs. Remember, the claim is not that there is a faculty for *language* as conceived of pre-theoretically – there is instead a faculty for *I-language*. As we have seen, commonsense categories are checked by science. In other domains as well, including common sense

³⁷ This is a wide-ranging issue; putting briefly what is relevant here, Dupré thinks there is no one such thing as “science”; there are instead a variety of ways to address different problems according to the demands posed by each one. With too simplistic an idea of science as a unified enterprise, he argues, one ends up with naïve answers to questions like “what is human nature?”. While I agree that “what is human nature?” is a question that is complex (i.e. divisible into parts), I do think there is something that typifies a scientific approach, even if what it amounts to is quite weak constraints about simplicity, elegance, empirical testing, etc. So we *can* say whether some answer to the question “what is human nature?” counts as scientific; but we should also recognise that there may be aspects to that question which will not be addressed satisfactorily by such an answer, and, conversely, that there can be ways of addressing the problem which may be informative and insightful yet not scientific.

itself, we can expect similar divergences between what it is that faculties underwrite, and how what they underwrite is thought and talked about. Therefore we should not necessarily be looking for LTs for abilities as conceived of in commonsense terms; and we should be particularly cautious about assigning LTs for complex behaviours, even if such behaviours can be given a relatively simple commonsense description.³⁸

Avoiding hasty steps from complex behaviour to unitary explanation is a problem that evolutionary psychology sometimes faces. Evolutionary psychology is a discipline which has benefited from increasing interest and allocation of resources over the last few years. The basic idea is that the mind can be seen as an “adaptive toolbox”, with different components that helped their possessors to tackle the problems encountered by man in a stone-age environment – the time frame of natural selection is such that this is the period which it is assumed the skills we have are a response to.³⁹ The approach is potentially an extremely powerful one. Undoubtedly in some areas where it is harnessed and carefully applied it may cast light on puzzling questions. However, a concern expressed thirty-five years ago is still extremely current:

[T]here has been a natural but unfortunate tendency to ‘extrapolate,’ from the thimbleful of knowledge that has been attained in careful experimental work and rigorous data-processing, to issues of much wider significance and of great social concern. This is a serious matter. The experts have the responsibility of making clear the actual limits of their understanding and of the results they have so far achieved, and a careful analysis of these limits will demonstrate, I believe, that in virtually every domain of the social and behavioural sciences the results achieved to date will not support such ‘extrapolation.’

(Chomsky 1968, p. iii)

³⁸ As indicated by Kaplan and Pigliucci (2001), there are difficulties associated with the expression “a gene ‘for’ X”. Similarly “LTs ‘for’ Y” may be problematic. Further problems could arise if LTs are equated with strict templates for modules.

³⁹ Interesting anthologies of ideas and positions in this field include, for example, Hirschfeld & Gelman (1994), Barkow, Cosmides & Tooby (1992), Carruthers & Chamberlain (2000).

Sometimes, extremely complex social behaviours, construed according to the terms in which they are thought of pre-theoretically, are explained by appeal to a component of the brain with a genetic basis. The component is often described as a module, with different levels of care taken, depending on the individual proponent of the claim, as to what properties this mechanism might have and the extent to which it is entirely pre-determined or interacts with environments. The presence of the module and its precise function are determined and justified by appeal to 'evolution'.⁴⁰ This can have the effect of excusing (or condemning) as inevitable behaviours which are in fact not necessarily entirely automatic.⁴¹ The positing of cognitive modules with a genetic basis, seen as evolutionary adaptations to problems faced in past environments, in order to account for complex current behaviours, exemplifies the need for ethnoscientific ideas and methodologies to be applied with clarity and accuracy, and for an awareness of the modesty of empirical evidence for such claims.

3.7 ETHNOSCIENCE IN THE CONTEXT OF WIDER SCIENTIFIC INQUIRY

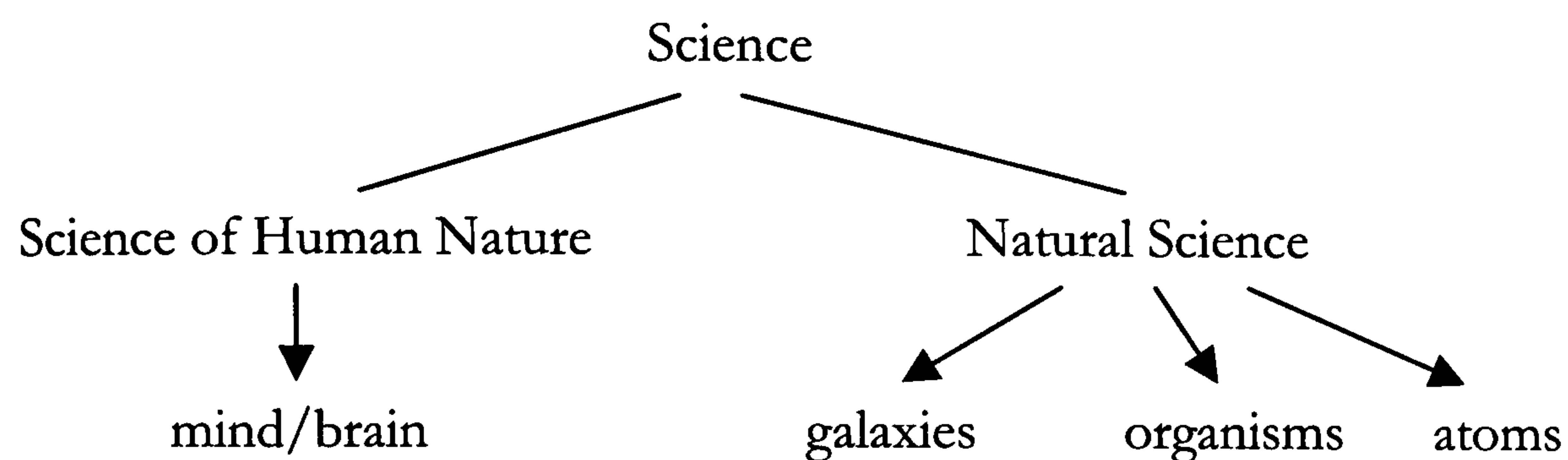
Before turning to concentrate on the study of commonsense psychology in the next chapter, let us set out where ethnoscience sits in relation to other scientific activities. In order to explore this question, I will rely on a series of annotated diagrams of the kind introduced in the first chapter, some of which will be repeated here alongside others which are developed and commented on in more detail. In the following figures, simple (non-arrowed) lines represent branches of sciences, while arrows show what some science is directed at as its main concern. It should be noted that the central aim is to identify topics of investigation and how they are related. Concerns

⁴⁰ See Fodor (2000b, chapter 5) for a critique of some of the excesses of evolutionary psychology. Fodor (1998a, chapter 17) provides a map of how nativism, modularity and Darwinism are and are not interconnected. Lewontin (1990) discusses evolutionary reasoning about cognitive processes.

⁴¹ See Dupré (2001, p.81-92) for some considerations about the ethical implications of socio-biology and possible repercussions on social policy; see also Kaplan (2000); Stich & Murphy (2000).

about what some science should be called or what topics properly fall under the jurisdiction of one science rather than another are only of secondary interest.⁴² Of course the diagrams are in no way supposed to be exhaustive; the sciences and topics that appear in the figures serve as examples, chosen mainly for illustrative or contrastive effect.

Fig. (3i)

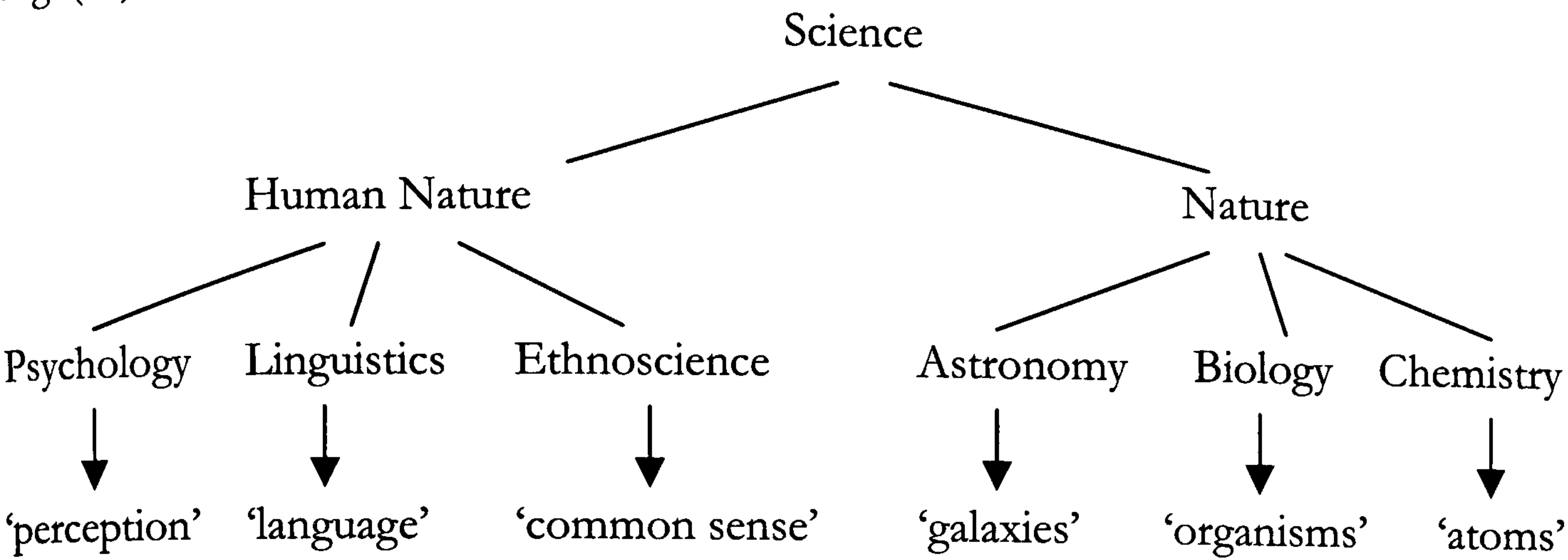


As has been emphasised from the outset, the study of the mind/brain, the science of human nature, falls firmly under the umbrella of natural science. Human minds are just another element in the natural world that we might set out to try to understand.⁴³ Different sciences study the various aspects of the natural world; in the same way, different sciences study various aspects of the mind/brain. We can separate some of these out as in figure (3ii):

⁴² An issue that does deserve attention is how existing terminology maps onto the one I have adopted. Atran (1990) approaches folk biology as I am advocating that folk psychology should be approached. In doing so, he introduces the terms ‘folkbiology’ and ‘ethnobiology’ (see also Atran 1998; Medin & Atran 1999; Wilson & Keil 1999). However, his ‘folkbiology’ is my *ethnoscience of biology* (i.e. naturalistic inquiry into commonsense biological understanding); his ‘ethnobiology’ is more like my *reflective commonsense biology*. See also Nuti (1999) for discussion.

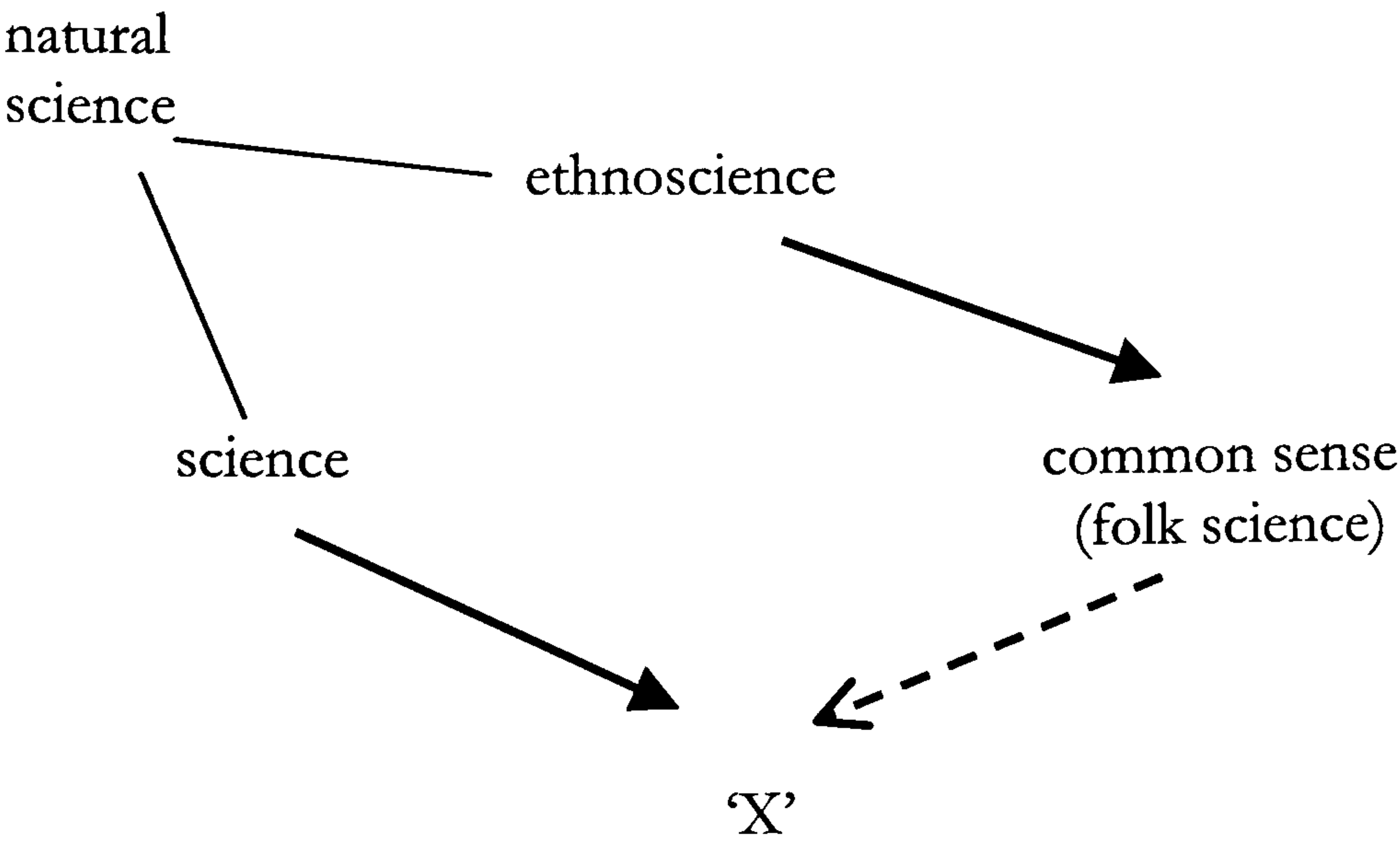
⁴³ The distinction is therefore rather arbitrary but useful for illustrative purposes.

Fig. (3ii)



Of course, in some sense linguistics and ethnoscience count as psychology, as all of the science of human nature could be argued to be, but it is helpful to make traditional boundaries a little more fine-grained as we consider particular specialisations. Let us now concentrate specifically on ethnoscience, which studies those aspects of the mind responsible for commonsense understanding. Ethnoscience can also be subdivided further according to what common sense is taken to be examining. Indeed, the iterative nature of this process is emphasised by figure (3iii), a slightly modified version of figure (1i) that was used when ethnoscience was first introduced in chapter 1:

Fig. (3iii)



Whatever aspect of the world is attended to as the X in the above diagram, including of course mental aspects of the world, ethnoscience studies how that aspect of the world is conceived of in terms of common sense, as shown in figures (3iv) and (3v):

Fig. (3iv)

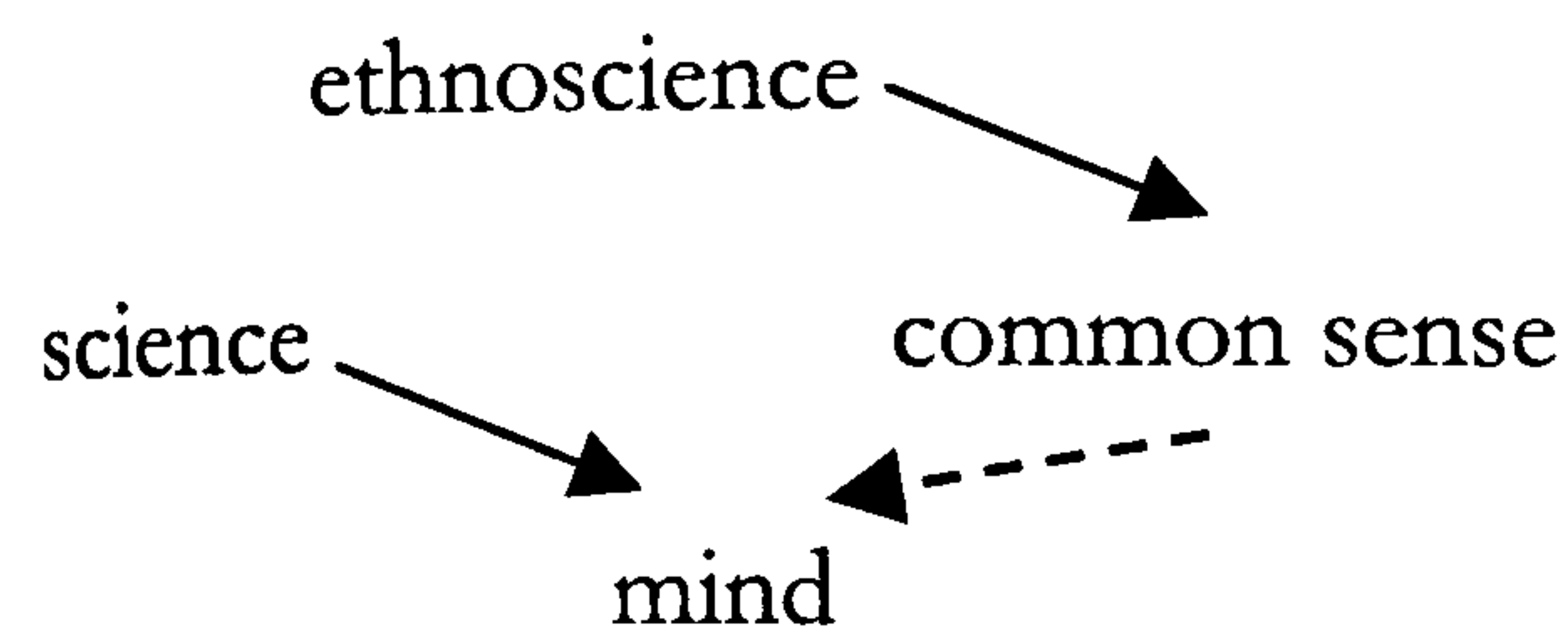
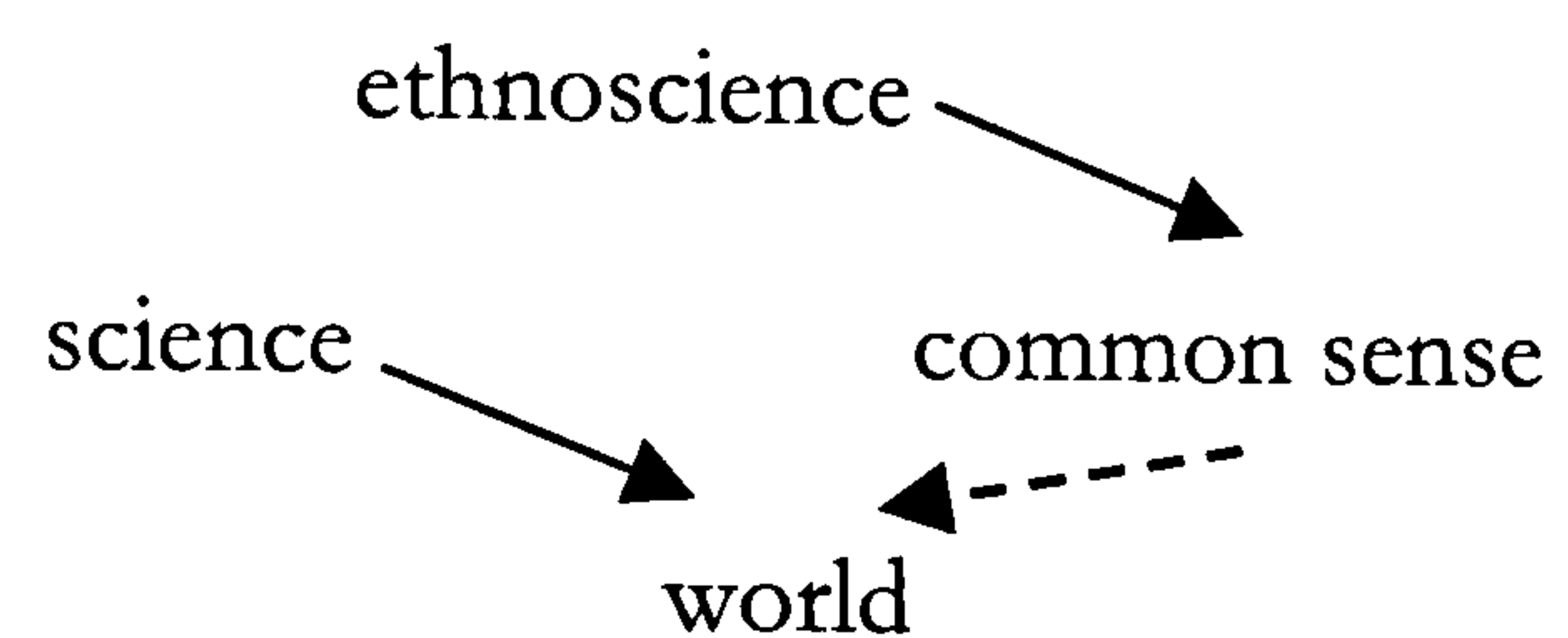
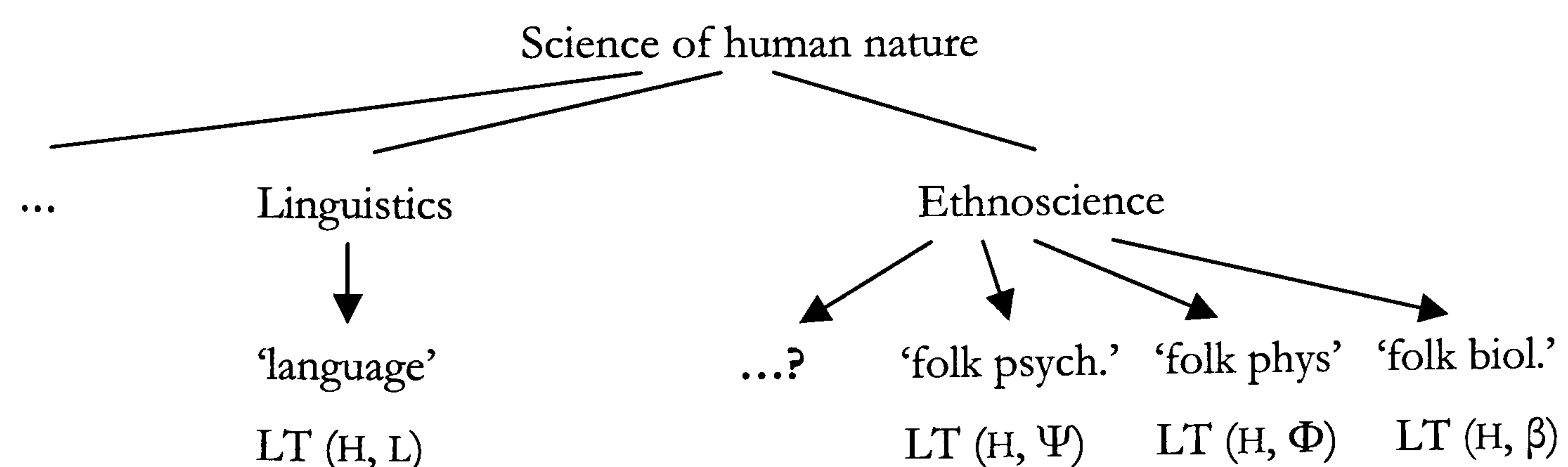


Fig. (3v)



At the stage this type of diagram was introduced, we were contrasting a scientific perspective with a commonsense perspective using different types of arrows (full and dotted respectively). As figure (3iii) emphasises, both ethnoscience and science are part of natural science. We are now in a position to be more specific about where ethnoscience fits in to the wider scheme of things and can return to the sequence started in figures (3i) and (3ii). Figure (3vi) follows on from these to illustrate how ethnoscience is often divided up:

Fig. (3vi)

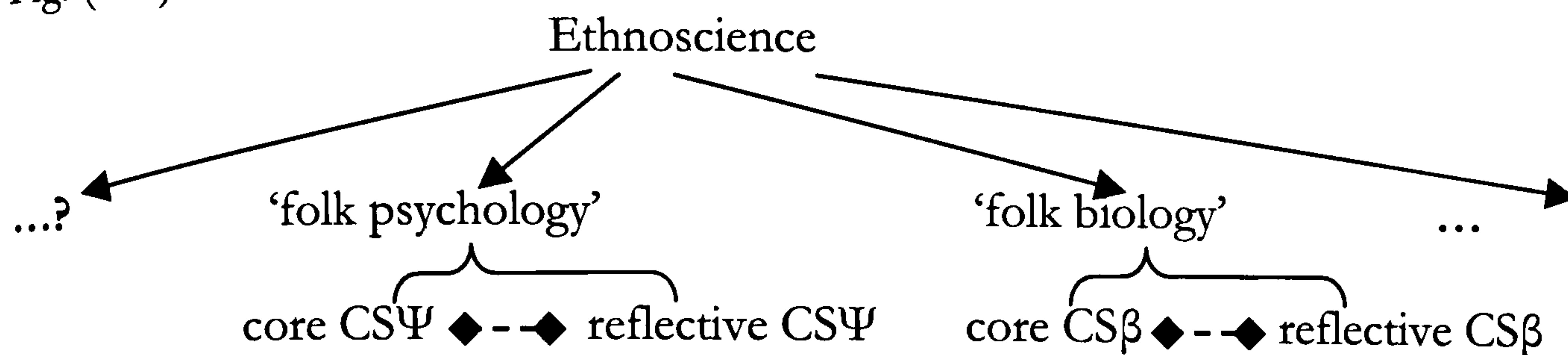


Looking at the diagram from the top going downwards, there is an extra branch to the left of linguistics leading to dots because there is clearly more to the science of human nature than just linguistics and ethnoscience, but these are what we are

discussing. That linguistics and ethnosience are on the same level serves to remind us that they study the two “cognitive structures”, grammar and common sense, that it was suggested be attributed to the organism at the outset. At the next step down, we find language at the same level as individual folk theories. So far, all the objects of study have been in quotes. This is to indicate that up until now they have been identified pre-theoretically. At the next level down again, we have the proper object of scientific inquiry, set out in the learning theory terminology. These placements highlight the fact that initial discussion of ‘folk theories’ mirrors initial discussion of ‘language’; and that eventually understanding of folk theories needs to be cast so that it mirrors, with relevant adjustments, the notion of I-language.

There are many aspects to what any of the above sciences study – galaxies, organisms, and atoms are only instances of what each of astronomy, biology and chemistry respectively aim to tackle, and are each terms covering a variety of referents; the same goes for the objects of inquiry of the ‘sciences of human nature’. As we focus on the science of human nature, we could divide up the subject a bit further. So linguistics breaks down into e.g. phonology, syntax and semantics. We might wonder whether any of the objects of inquiry of ethnosience justify similar sub-disciplines. Finally, just as linguistics had to place itself in relation to other approaches for its significance to be fully recognised (as discussed at the end of chapter 2), so it is part of the concern of ethnosience not just to characterise the abilities of folk science (and their acquisition) correctly, but also to relate that aim and its results to other notions of folk science and to early science. In practice, this process might contribute to the understanding of reflective common sense, introduced in section 1.3. As noted, we want to leave open the possibility that, to fully account for how common sense works, ethnosience will have to address this too. Figure (3vii) provides a reminder of where reflective common sense sits.

Fig. (3vii)



Here we can see that, for each of the areas into which common sense can be divided up (some of which are not shown here but have been mentioned, as indicated by the arrow to the dots on the right; some of which may not yet have been considered, as indicated by the dots on the left with a question mark, because how many areas common sense divides up into is an empirical question), ethnoscience will attempt to provide a theory of the contribution of the output of its LT and of the place of reflective common sense in that area (in other words, attempt to elucidate the nature of the ‘connection’ between core common sense and reflective common sense). A suggestion about how this might work, first for psychology and then for language, will be put forward in the next chapters.

3.8 TERMINOLOGICAL NOTE

At the beginning of this chapter we noted some reservations about the term *folk*. While it might seem desirable therefore to avoid it altogether and replace it with something else instead, we doubted whether it would be possible to get rid of *folk* entirely as it is widespread in existing literature and no doubt will continue to crop up. It is more important then to make sure that we be absolutely clear what we are talking about when we encounter it. For there is another, more serious, problem faced by most discussions about folk sciences, that goes beyond the connotation of terms and remains to be sorted out even if we substitute *folk* with *naïve* or *commonsense*. The

terminological differences we encounter arise not so much because the same things are talked about in different ways; when it comes to accounts of folk (or naïve, or commonsense) science, people are often talking about very different things with the same, or interchangeable, labels.

As an example (paraphrasing Chomsky), discussions of folk psychology are often conducted without basic conceptual distinctions being drawn or recognised. For instance, 'folk psychology' can refer to everything from (i) some analogue, for psychological understanding, to UG and its output, to (ii) the general interests of a particular educated Anglo-American culture area, to (iii) the refined technical philosophical discourse of the academic community. The problem is exacerbated by the fact that the area of investigation is generally individuated very broadly as being what provides an everyday ability to interpret others. And any of the information listed above (possibly not so much the last) may well be called upon in interpreting others, as may much else, given how wide a range of activity falls under this description.

In short, depending on how a domain is identified in the first place, many factors could be relevant to the abilities humans have in the area in question. Just because a competence is called upon in some task it does not follow that just *any* thing that is called upon in performing the task can be ascribed to the competence. Clearly there is a lot of other information about domains available to individuals that could be relevant to skills in some domain. Unfortunately all of this all too easily gets classed as folk knowledge, in some or other sense. To clarify, what you get from an LT (H, D) is core-commonsense understanding, or folk knowledge in a rather restricted sense of *folk*. This core-commonsense understanding may then be elaborated in certain ways. Some ways of re-elaborating core-commonsense understanding are more interesting for science than others: those where the re-elaboration appears to be a general human trait with variation across identifiable parameters.

Basically, I think there are two kinds of common sense that deserve to be the object of rationalist naturalistic inquiry into the mind/brain. On top of that, there are further elements which contribute to the skills and abilities characteristic of humans as they engage with the world in everyday activities in virtue of what might be labelled pre-theoretically as commonsense understanding. These latter, further elements are what I will be calling *encyclopaedic* \times (encyclopaedic psychology or language or whichever we want to look at). I do not discuss them in detail because, while they may well in part depend on the main two kinds of common sense in some cases, they are likely to be a lot less stable and universal. Also, it may be that these elements don't have to be reliant on the two main kinds of common sense – consider the high-functioning autistic's explicit conception of mind, presumably built up on the basis of general processes; or the body-language expert's explicit ideas about correlations between postures and attitudes, which stem more from ethological observations rather than natural endowment specific to the domain and normally applied; or a scientist's research results extended, however successfully, outside the laboratory.

The two kinds of common sense I want to focus on have not always been recognised as distinct, which leaves scope for confusion. The first, call it core-common sense, once properly formally characterized theoretically, may be counterintuitively distant from a recognizable everyday framework. It does however form the basis of what humans call upon in forming their representations and expectations of the world around them. The second kind of common sense provides humans with an awareness of what the core-commonsense expectations are in such a way as to be able to reflect on the structure of the world as interpreted for them by their core-common sense. There may be different ways of doing this, and to some extent how it is done may be dependent on natural language resources. I doubt, however, that it is an entirely idiosyncratic process, at least for some basic aspects of reality. To the extent the process is predictable and testable we are faced with a problem rather than a mystery.

4

LINGUISTICS AND THE ETHNOSCIENCE OF COMMONSENSE PSYCHOLOGY

4.1	THE OBJECT OF INQUIRY AND THE DOMAIN OF THE OBJECT OF INQUIRY.....	104
4.2	METHOD IN THE ETHNOSCIENCE OF COMMONSENSE PSYCHOLOGY.....	111
4.2.1	<i>(R)evolution from early science of commonsense psychology.....</i>	<i>112</i>
4.2.2	<i>Grice and naturalistic inquiry.....</i>	<i>114</i>
4.2.3	<i>Constructing theories.....</i>	<i>118</i>
4.2.4	<i>Towards mindreading.....</i>	<i>121</i>
4.3	AN OUTLINE.....	127
4.3.1	<i>Methodological preliminaries.....</i>	<i>129</i>
4.3.2	<i>UΨ, I-psychology and the Psychology Faculty.....</i>	<i>133</i>
4.4	ACTION PARSING	135
4.5	UNDERSTANDING BEHAVIOUR.....	141
4.5.1	<i>Levels of analysis and levels of understanding.....</i>	<i>142</i>
4.5.2	<i>Complex and basic behaviours and interpretations.....</i>	<i>144</i>
4.5.3	<i>The problem of over-attribution.....</i>	<i>145</i>
4.5.4	<i>A 're-interpretation' hypothesis.....</i>	<i>147</i>
4.6	OUTLINE REVISITED.....	152

4

LINGUISTICS AND THE ETHNOSCIENCE OF COMMONSENSE PSYCHOLOGY

I have argued that the study of language as pioneered by Chomsky provides a model for the study of common sense. We now turn in particular to the study of commonsense psychology. Making use of the concepts and distinctions discussed in previous chapters, the plan for the rest of this thesis centres on a constant comparison between language and mindreading, and by extension between linguistics (and its sub-disciplines) and accounts of commonsense psychology. For particular issues⁴⁴, the aim is to consider how the case of mindreading matches up with, or differs from, the case of language. The idea is that questions faced in the study of language are also likely to arise for commonsense psychology. Expecting similarities in the nature of what is being investigated, we might also find similarities with regard to how the issues have been debated. If considerations familiar from the study of language can be carried over to the area of mindreading, we might learn something not only about what kind of thing commonsense psychology is and how it works, but also about how problems might be approached and how to evaluate existing proposals. If the familiar considerations do not transpose, this strategy should still be useful to highlight differences between the two domains.

⁴⁴ For example, how the object of inquiry is conceived of pre-theoretically in each case (in this chapter), and comparisons with regard to truth and accessibility of knowledge (in the next chapter).

4.1 THE OBJECT OF INQUIRY AND THE DOMAIN OF THE OBJECT OF INQUIRY

Let's pretend there's a Psychology Faculty (FP), a component of the mind in the same spirit as the Language Faculty (FL). The Psychology Faculty⁴⁵ enables humans to make judgements about behaviour in terms of underlying unobservable psychological states. That we are interested in a *Psychology Faculty* should already suggest the kind of approach that aims to establish the cognitive capacities of humans in a particular area rather than attempting merely to systematise external manifestations of those capacities. In other words, we could say we will be investigating 'I-psychology', echoing the introduction of 'I-language'.

Stich & Ravenscroft draw a distinction along these lines between 'external' and 'internal' accounts of folk psychology. They note how on external readings, the label 'folk psychology' could refer to "the collection of folk psychological 'platitudes' that people in our culture readily recognize and assent to". Or it might refer to "a set of more abstract generalizations – a 'theory' if you will – that systematizes the platitudes in a perspicuous way", noting that such "systematization might well invoke terms and concepts that are quite unfamiliar to ordinary folk" (Stich & Ravenscroft 1996, p. 127-128). What both these 'external' notions have in common is that in neither case do we expect that the 'theory' is necessarily what underlies the abilities exhibited in producing the intuitions. On 'internal' readings, 'folk psychology' is reserved for internally represented psychological rules and generalizations that underlie so-called folk-psychological abilities.

It is certainly important to be clear about whether we are constructing an internal or an external account. Failing to do so can lead to the kind of disagreements that famously occur between linguists and philosophers of language.⁴⁶ The

⁴⁵ Segal (1998) also uses this terminology, suggesting that the Psychology Faculty uses representations made available by the Language Faculty. The idea of a Psychology Faculty is developed in a different way in what follows below.

⁴⁶ For instance, as noted in chapter 2, on the problem of extensionally equivalent theories; see e.g. Quine (1960) and Chomsky (1969, p. 272-273; 1986, p. 245).

introduction of the notion of I-language marks the shift of focus onto a component of the mind/brain rather than sets of sentences. However, this is not the only purpose of the introduction of the notion of I-language. Talking about *I-language* also serves to distinguish between phenomena related to language pre-theoretically, and phenomena related to the object of scientific inquiry. A parallel ground-clearing move needs to take place for commonsense psychology too. We don't just want to shift blindly to an internal account – in doing so we should be aware that we might have to circumscribe what it is that such an account might plausibly explain.⁴⁷

In order to study the Psychology Faculty, we should first consider what kind of facts we expect it to underwrite. For example, as discussed in chapters 1 and 2, in the case of language, the commonsense idea of 'language' – or some early science development of that notion – is not what we expect naturalistic linguistics to take as its object of study. For instance, that there are many different languages which match up with state boundaries is not a fact that an internal account of the psychology of an individual can or needs to address. Are there any comparable phenomena that the Psychology Faculty might erroneously be expected to account for? The initial answer appears to be 'no', perhaps because it is strikingly difficult to reify commonsense psychological understanding. Support for this observation is found readily in the everyday world. If you tell someone that you do linguistics, usually you have to spend some time explaining why you may not be multi-lingual. If you say you work on commonsense psychology, you have to explain what that is in the first place. People might hazard a guess, but probably because they attempt to make some sense of the label rather than because they have a readily available notion they latch onto.⁴⁸

⁴⁷ As we shall see, the challenge in identifying the domain of I-psychology stems not so much from interference from everyday counterparts to the notion of commonsense psychology, which turn out to be scarce. The difficulty is more that some of the questions that accounts of commonsense psychology are widely expected to address are chosen with an eye on pressing problems in the philosophy of mind rather than on establishing the nature of a Psychology Faculty.

⁴⁸ Unaided they tend to put forward slightly the wrong idea; with help they tend to focus on particular instances of interpretations of behaviour, actual predictions and explanations. As with

We can speculate as to why this might be. For instance, we might think that while some of the results of the Language Faculty are tangible, the Psychology Faculty deals with more ephemeral matters. Clearly, you can hear words and sentences; you can't see thoughts, or analyses of behaviour.⁴⁹ Nevertheless, it seems strange that something as powerful and pervasive as commonsense psychology has no immediate pre-theoretic counterpart (in marked contrast with the case of language). Possible candidates might be *behaviour* or *mind*. However, although clearly related to commonsense psychology, neither seems quite right. *Behaviour* is indeed a bit like the commonsense notion of *language*, focusing as it does on external productions; but the point about commonsense psychology is that it concerns the *understanding* of behaviour, not just behaviour *tout court* (although of course understanding behaviour is an instance of behaviour). And while *minds* may be associated pre-theoretically with ideas about causing behaviour, that they also interpret behaviour is hardly even considered. Perhaps the more something is all-encompassing the more it is taken for granted. People read other minds so constantly and so consistently that they don't appreciate that what they are doing is quite a remarkable feat. Just as linguistics ends up bringing to light what people don't know they know about language, so an account of commonsense psychology could

language, then, instantiations of the application of the abilities rather than their underlying explanation are most salient. Eventually, though, people are usually less surprised that you study some aspect of the mind than they sometimes are when you explain what linguistics is actually about.

⁴⁹ This intuition, that the FL has a clear product (utterances) but the FP doesn't ('understandings?'), is actually quite misleading. Of course people speak, they write things down, they experience variations in sound-meaning pairs both in their own and foreign environments; however, not all of the results of the Language Faculty are tangible in this way. The structural descriptions and the associated interpretations assigned to utterances by the Language Faculty are no more tangible than the 'understandings' produced by the Psychology Faculty. What makes the language case feel more tangible is the Language Faculty having a perceptual/articulatory system attached to it. We might begin to wonder about positing an equivalent associated perceptual/articulatory system attached to a Psychology Faculty; e.g. possibly for the expression or portrayal of emotions, or the acting out of one's beliefs and desires.

discover what people don't realise they know about mindreading. That is, in other words, what people don't realise they know about how they *understand* behaviour; not just what people don't realise they know about behaviour.

There are of course everyday ideas about some bits and pieces that are connected to commonsense psychology. We have, for example, concepts like *thought* and *intention* and *reason* and *action* just as for language we have concepts like *word* and *sentence*. Indeed we seem to have rather more 'vernacular-commonsense'⁵⁰ concepts about psychological matters than we do about linguistic matters. This is hardly surprising: language may be more perceptually salient, but reflecting on commonsense psychology is more intellectually rewarding. Instances of psychological understanding are something we consider more and in turn talk about more than linguistic ability. But, I shall be arguing, there is no more reason to assume that what we are able to think and talk about when it comes to commonsense psychology exhausts what there is to know about our psychology faculty than there is reason to believe that all there is to linguistic ability can be introspected (see 5.2.2).

So, despite the fact that it is possible to latch onto the general area that theorists are interested in investigating more or less immediately – it is an area that is easy to think about, intuitive and appealing, where even brief and superficial introspection appears to yield insights worth defending – unlike 'language', 'commonsense psychology' *as a whole* is not something people generally think about. Note that it is necessary to introduce new phrases and expressions to refer to it (in any language, so far as I am aware); hence in English, 'commonsense psychology', 'folk psychology', 'theory of mind'. These compound neologisms are not, at least initially, terms of art; they are needed to label an area of interest even before any motivation has been established

⁵⁰ I am introducing the term 'vernacular-commonsense' to avoid possible confusion between the idea of a commonsense concept understood as a concept of commonsense understanding (whatever such concepts may turn out to be), and vernacular-commonsense concepts, i.e. everyday notions. The label is borrowed from Griffiths (1997, p. 200-201). See also 4.6 and chapter 5 below.

for a specific term with particular theoretical commitments. The area of interest thus picked out tends to be glossed, pre-theoretically, along the lines of ‘predicting and explaining behaviour’ and illustrated with a variety of examples.⁵¹ Unfortunately, the very introduction of new terms leads to the impression that a certain amount of theoretical work has already taken place before their introduction, when in fact in this area that is not necessarily the case.⁵²

So, when what is wanted is a way of picking out the abilities and practices at issue, I think *mindreading* is a better choice. For a start, ‘folk psychology’ is misleading (see chapter 3 on the connotations of ‘folk’); and, according to some, ‘theory of mind’ is presumptuous (what if theory of mind turns out to be not a theory after all? It might turn out to be a simulation process instead; or, more likely, it could be theory-like yet not be constructed exactly like an explicit scientific theory).⁵³ Going beyond questions of style, and more importantly, these terms are used on occasion not only to describe particular abilities but also to refer to what is thought to underlie those abilities. Such metonymic usage must be handled carefully. Identifying the practices

⁵¹ Practices at issue in most discussions of commonsense psychology include the ability to attribute beliefs to other people on the basis of their surrounding environment and other factors and draw inferences about further beliefs that can be attributed to them; the ability to explain and predict behaviour on the basis of the attribution of both true and false beliefs (and desires); the ability to produce descriptions of behaviour using mental state terms; the ability to recognise psychological generalizations as pertaining to instances of behaviour; understanding the meaning of mental state terms like ‘belief’ and ‘desire’ (see e.g. Stich & Nichols 1996, p. 138).

⁵² As Stich & Ravenscroft acknowledge, “in the absence of a theory about the mechanisms underlying these abilities, there is no obvious or natural way of drawing a boundary and saying exactly which capacities do and do not count as folk psychological capacities. That vagueness will make for problems as we proceed” (1996, p. 126).

⁵³ There has been much debate between *theory theory* and *simulation theory* accounts of variously circumscribed ‘folk-psychological abilities’, and between different versions of simulation and theory accounts. See Davies & Stone (1995a) and (1995b), and Carruthers & Smith (1996) for collections of papers on these issues. It is part of what I want to argue that not following the methodological practices presented here has led this debate to meander; see 4.6 and 5.2.2 below.

and the mechanism⁵⁴ by the same label isn't of itself problematic. The problems arise if you expect or require that the mechanism so-labelled must account for all the practices initially (and pre-theoretically) singled out for investigation.

Identifying what counts as commonsense-psychological *intuitively* is fine for homing in on an area for study. But precisely one of the things that science might hope to discover is whether some cluster of phenomena grouped together by common sense in fact share a common explanation. While 'predicting and explaining behaviour' is a perfectly satisfactory way of describing what is meant by 'mindreading', we should not presume that anything and everything that might be called upon in predicting and explaining behaviour should be attributed to 'commonsense psychology', 'folk psychology', or 'theory of mind' *when these are construed as the label for whatever it is that underwrites humans' ability to understand behaviour*. Hoping to avoid potential confusion, I therefore tend to use 'mindreading' for certain abilities and practices that form the initial basis of our interest, and reserve 'commonsense psychology' for something closer to I-psychology.

The prediction and explanation of behaviour could, on occasion, call upon practically any piece of knowledge or information. It is unlikely that everything that is appealed to in mindreading is directly the result of the Psychology Faculty, although a lot of information may be fed into or passed through the Psychology Faculty in order to come up with a final result. It is particularly important to bear in mind this point in discussions about whether commonsense psychology is modular, or whether commonsense psychological knowledge is tacit knowledge.⁵⁵

To draw a comparison, it is not the case that everything involving language or linguistic ability is ascribed to the Language Faculty. Specifically, it is a common error

⁵⁴ I am using 'mechanism' here loosely to mean underlying explanation, without commitment to any specific structures.

⁵⁵ See for example Heal (1996) and Gordon (1995) for discussions of modularity and tacit knowledge of commonsense psychology where this issue is not taken into account.

to conflate language with communication. The error is compelling because communication is perhaps the activity with which language is most closely associated. But while natural language can be used to communicate, communication is not the only purpose language serves, nor is natural language necessary for communication to take place (see Sperber & Wilson 1986/95, chapter 4). Recognising this fact frees up both accounts of language and accounts of communication, and gives rise to correspondingly separate branches of linguistics. The possibility of different aspects of mindreading abilities being addressed by sub-disciplines in the study of commonsense psychology may be worth considering. Otherwise, approaching commonsense psychology as a unified whole on the basis of a pre-theoretic gloss means that we risk studying commonsense psychology as if we were trying to do phonetics, phonology, syntax, semantics and pragmatics all rolled into one.

Note that I am not assuming that the Psychology Faculty and the Language Faculty will turn out to be exactly the same kind of system, but the idea is to study them in the same way. So I am not suggesting we look for exact counterparts to, say, phonetics and phonology, but that just as there are sub-domains in linguistics, we should check whether there are for the study of commonsense psychology. Speculating as to what they might be could lead to fresh insights. There may be different components within the Psychology Faculty, responsible for different aspects of mindreading.⁵⁶ Or some aspects of ‘mindreading abilities’ might be best accounted for from outside the Psychology Faculty.

So, not everything that is to do with ‘psychology in some sense’ or ‘what is called upon in understanding behaviour’, is necessarily the proper object of investigation for a naturalistic study of commonsense psychology. Of the things that are, not all of them will receive the same kind of explanation. Compare: not all of

⁵⁶ Examples could include Baron-Cohen’s Shared Attention Mechanism (SAM) and Eye Direction Detector (EDD); Leslie’s Theory of Mind Mechanism (ToMM) and Selection Processor (SP). See e.g. Baron-Cohen (1994, 1995); Baron-Cohen & Swettenham (1996); Leslie (1987, 1994a, 1994b, 2000); Leslie & Polizzi (1998).

what counts as language pre-theoretically is of concern for scientific accounts of language; of the problems that are relevant, some might be explained by syntax and others by pragmatics. Similarly, when we humans engage in behaviour-explaining behaviour, i.e. the behaviour that is mediated by (among other things) our commonsense psychology, we may call upon all kinds of information (e.g. encyclopaedic) and all kinds of abilities (e.g. logical and metarepresentational), only some of which can be directly related back to commonsense psychology. Of the things that are commonsense-psychological, some could receive the equivalent of syntactic explanation, some might receive the equivalent of a pragmatic explanation.

In section 4.3, I provide an outline of possible components contributing to a Psychology Faculty and in the rest of the chapter I discuss them in the context of current research. But first, we need to consider general constraints on how to proceed. In particular, we need to adopt a formal methodology to introduce and define our technical terms and according to which success or failure of a proposed theory can be evaluated. As it turns out, some very interesting suggestions relevant to the development of ethnoscience of commonsense psychology have been made by a philosopher.

4.2 METHOD IN THE ETHNOSCIENCE OF COMMONSENSE PSYCHOLOGY

Naturalistic inquiry into commonsense psychology is inchoate. As one starting point, we might look to the ideas of early science of the domain we are interested in. After all, Chomskyan theories of syntax initially borrowed and adapted the terminology and generalizations of grammarians and philologists; perhaps philosophical inquiry into commonsense psychology can help us in the same way. There is a sense in which certain kinds of philosophy of mind stand to our endeavour as traditional grammar stands to the first theories of generative linguistics. And just as some traditional

grammarians⁵⁷ produced work which was more in keeping with early generative syntax and hence of most use to linguists, so some ideas from philosophy of mind are more congenial to our purposes than others.

4.2.1 (R)evolution from early science of commonsense psychology

Bearing in mind these observations, we examine here in detail the ideas presented in Grice's paper *Method in Philosophical Psychology: From the Banal to the Bizarre* (1975).⁵⁸ I think that its proposals deserve more attention than they have been given until now; the paper is strikingly current and makes suggestions that can be adapted to the methodological approach we are pursuing. That Grice approaches commonsense psychology from a philosophical perspective which is compatible with the concerns of psychologists of common sense (ethnoscience) and cognitive ethologists is of inherent interest. In what follows I intend to draw freely on *MPP* to illustrate my ideas. We should be aware, however, that Grice's paper is ultimately in philosophy of mind, and that at certain points our interests and those of Grice diverge.

Grice seeks to analyse psychological concepts like *belief* and *desire*. He notes that attempts to provide explicit definitions quickly encounter problems of circularity, which, he suggests, may be overcome by defining these notions implicitly via a theory. The paper contains proposals for how a psychological theory suitable to this purpose might be devised. Grice further aims to show how the provision of such a theory is related to other issues surrounding these "central psychological concepts" (1975a, p. 122), for example the status of psychological laws as necessary or contingent and the question of privileged access to mental states.

⁵⁷ Jespersen, for example; see Chomsky (1977, chapter 1).

⁵⁸ First delivered as a Presidential Address to the American Philosophical Association, this is a little known, or at least little discussed paper – perhaps because it does not appear in the most familiar of Grice's collections (Grice 1989). Henceforth *MPP*, the acronymic potential of the paper's title may not have been wholly unintentional as the paper contains a hypothetical cognitive evolution whereby creatures come to be able to perform logical inference.

MPP is thus aimed at resolving specific philosophical questions. However, there are many reasons why it is of interest for the development of the ethnoscience of commonsense psychology. For instance, while Lewis's *Psychophysical and theoretical identifications* (1972) is a precursor of *MPP* and rather better known, Grice's proposals about the content of commonsense psychology are far more detailed and considered. Some features of these proposals are adopted in my outline of FP in later sections. Grice's view is also more consonant with methodological naturalism. Of particular value is that external accounts of commonsense psychology are rejected in favour of internal accounts, and that the theory of commonsense psychology is cast in technical theoretical terms. Grice's proposal also embodies an ethological realism which I wish to endorse. Finally, Grice adopts the description of behaviour in a mentally loaded way, i.e. as action. This is in tune with the internalist theory of I-psychology envisaged here, even though *MPP* provides us with little explicit justification for it.

Just before considering Grice's concrete proposals, there is a further point to make about probable consequences of our diverging aims. Acknowledging differences between naturalistic inquiry into I-psychology and Grice's project is important both in its own right, and because the situation here mirrors that noted above (at the end of chapter 2) for the study of language. Grice supposes that everyday speech and thought about psychological matters is underwritten by a psychological theory, and he is well aware that more than one theory might fit the bill. A major concern in *MPP* is to show how choices about the type of theory constructed would have an effect on the resolution of core philosophical problems. Thus each of these, and Grice's desire to honour particular intuitions about the mind, imposes constraints on the type of theory we consider. However, we might wonder whether an account of commonsense psychology, understood as I-psychology, will answer a pre-determined set of philosophical problems; in much the same way as we doubted that linguistic theories about semantics need satisfy all that might count as the requirements of philosophical theories of meaning (see section 2.4.2 above).

For instance, while agreeing that in some sense everyday speech and thought about psychological matters may be underwritten by a theory (or theories, see chapter 3), we might not expect everyday notions or terms (vernacular-commonsense psychology) to be analysable in any straightforward way by reference to the concepts of what we might think of as core-commonsense psychological theory, or I-psychology. Similarly, we cannot expect in advance that philosophical questions such as those about privileged access to beliefs will be resolved by our approach, although we may hope that an account of the Psychology Faculty could say something about these matters. Having acknowledged these reservations, we can now turn to consider the parts of *MPP* which are most relevant to our concerns.

4.2.2 *Grice and naturalistic inquiry*

The general strategy in *MPP* is to construct psychological theories to provide explanatory bridges, for different creatures, between being in certain situations and engaging in particular behaviours. The psychological theories will vary according to the scope of the behaviours the theory has to accommodate. Each theory involves psychological concepts defined in terms of the laws, or quasi-laws, of the theory. The overall hypothetical programme for philosophical psychology involves a series of nested theories where the terms of the more complex theories are inherited from the less complex, where possible; eventually a theory of the level of complexity required to account for human behaviours, including social behaviours, is reached.

While Grice recognises that this sort of approach (explicating psychological concepts by relating them to laws of a theory that systematically generates explanations of behaviour) may not be new, he intends to take it further than others:

What is needed is rather more attention to detail than is usually offered by advocates of this approach; we need to pursue such questions as what the special features of a psychological theory are, why such a theory should be needed, what sort of laws or generalities it should contain, and precisely how such laws may be used to explicate familiar psychological terms.

(Grice 1975a, p. 125)

So, Grice aims to develop a concrete and detailed proposal within an entire framework (“My hope is, in effect, to sketch a whole system”, *ibid.* p. 121), making explicit and defending his assumptions (including appeals to evolutionary considerations). The procedure he outlines can be applied to different sorts of creature and is sensitive to variation between creatures. The complexity of the required psychological theory will depend on the complexity of the manifested behaviours that the theory for some creature has to account for. In implementing this proposal, care is taken with determining the features to attribute to a psychological theory of this kind, the specification of the actual laws and generalizations that the theory is to contain, the reasons for requiring the theory and the relation of familiar psychological terms to the theoretical concepts introduced.⁵⁹

MPP thus has many features of interest for naturalistic inquiry into commonsense psychology. The introduction of technical concepts like, for example, JUDGING, WILLING, PREHENDING (rough counterparts to *believing*, *wanting*, and *perceiving*, as we shall see in the next sub-section) helps to avoid pre-theoretic discourse and understanding influencing investigation into the underlying nature of commonsense concepts. Furthermore, it is possible that technical terms could be introduced in a project of just systematising linguistic intuitions about statements concerning psychological matters;⁶⁰ however Grice rejects this option in favour of one that “might in the end tell me that a particular theory is the right one, by some test beyond the ease and elegance with which it fits the data” (*ibid.*, p. 139). This decision indicates a shift of focus from external to internal accounts, and a desire to go beyond descriptive adequacy towards explanatory adequacy.

⁵⁹ As will be discussed in the next chapter, Grice’s view of the task of explicating commonsense psychological concepts is in marked contrast to that of many interested in this area of philosophy of mind (see 5.2.2). The standard view seems to be that the laws of commonsense psychology are largely self-evident and can be adduced with a moment’s reflection.

⁶⁰ Recall from the beginning of this chapter that such a process would give rise to an ‘external’ notion of folk psychology in Stich & Ravenscroft’s terms.

Another reason Grice dismisses the intuition-systematizing option for arriving at a psychological theory is that it does not comply with his desire to capture the continuity that he sees as holding between different psychological theories ascribable to different creatures. “I am much impressed by the fact that arrays of psychological concepts, of differing degrees of richness, are applicable to creatures of differing degrees of complexity, with human beings (so far) at the peak” (ibid., p. 140). The intuition-systematizing option cannot do justice to this idea of continuity, or even to the idea of creatures other than humans possessing mental states of any kind. It presupposes mentalizing and presupposes language and language systems complex enough to talk about mental states. It is only applicable insofar as the creatures about which the theory is being constructed have (a) reflective access to and/or awareness of their mental states and (b) language to report their intuitions about them. The option is thus inherently inappropriate for investigating psychological theories of creatures other than humans.⁶¹ The main point about continuity I want to underline is the assumption that there exist mental states in nature other than in humans; for it is presumably not the case that there is a sudden jump from no mental states in nearly all animals, with humans alone on the other side of the divide (Allen & Bekoff 1997). The question is what kinds of mental states are to be found in different creatures, and Grice illustrates one way of investigating them.

Grice’s suggested method is to imagine inventing a series of creatures (*pirots*, varieties of which correspond to different real world creatures) that have to solve the same kinds of problems as their existing counterparts. “The general idea is to develop sequentially the psychological theory for different brands of pirot” (1975a., p. 140). In this picture, we pretend to be the *genitor* faced with the problem of designing

⁶¹ This is a relevant point for cognitive ethologists for whom it is important that how we talk about ‘beliefs’ and ‘desires’ should not prejudice issues about the existence and nature of animal mental states. On language use and attribution of belief, see Allen (1992); Kornblith (2002, p. 83-89).

operants (living things).⁶² Each operant has an energy store, and performs a certain set of operations that use up energy but that are necessary to keep the operant functional. In fact, operant-types could be defined by what the set of such operations is. The genitor is concerned to optimize survival chances:

if an operant (x) is not survival oriented, there is no basis for supposing it to exist at all; since [...] x will not exist unless pro-genitors are around to produce it; and if they are to have the staying power (and other endowments) required for x 's production, x , being of the same type, must be given the same attributes. So in providing for the individual x , some provision for the continuation of the type is implicit.

(ibid., p. 141)

So, we consider sequential brands of pirot with sequentially more complex psychological theories depending on the complexity of the minimum set of operations that each brand of pirot has to carry out in order to survive in a given living condition. For each pirot, we might ask what kind of psychological apparatus needs to be assigned to it in order that it should be able to perform its operations. For example, we can explain how plants perform their operations without invoking any psychological apparatus. Within the class of pirots for which it will be necessary to postulate psychological apparatus in order to explain how it is that they perform their operations, there should be an increasing order of complexity to this psychological apparatus depending on how demanding is their living condition. Living conditions could vary from the operant not having to move for its energy source, to needing to locate sources of energy, to dealing with different and changing environments, and dealing with other operants. A further variable would be how effective an operant has to be in its living condition.⁶³

⁶² Grice here talks of *operants* possibly because he is concerned with constraints on real living things. *Pirots*, by contrast, are models of actual creatures.

⁶³ There are some similarities between this enterprise and a project that aims to develop a computer programme (*Chimpworld*) that simulates living conditions and interactions of chimpanzees to establish

Imagining a procedure of this type brings into focus a variety of questions of detail.⁶⁴ We can entertain the possibility of different levels of JUDGING for different kinds of creature, and perhaps even different relations between psychological states for different creatures. Also we might distinguish more than one type of JUDGING within the same creature. For instance, in humans what is recognized pre-theoretically as a ‘belief’ state might cover more than one kind of JUDGING. Let us now examine how the laws of our theories might be framed.

4.2.3 *Constructing theories*

According to Grice, the psychological theories we aim to construct are to contain *ceteris paribus* laws and statements with the property of ‘law-allusiveness’. The feature of law-allusiveness allows us to state that there is such and such a law without being able to specify it fully, and thus covers cases where we have a law-like correlation between two variable functions at least one of which we currently cannot measure. For example, the statements ‘*An organism left for a long time without food will want food*’ and ‘*The longer an organism is left without food the more it will want food*’ have the feature of law-allusiveness. It will be the task of a more developed theory to produce laws. The *ceteris paribus* laws will be *ceteris paribus* for two reasons. First because *ceteris paribus* laws can be modified without being amended: they can be specified in more detail without otherwise changing their content. So approximations of the laws can stand in for the real ones while we work out what the real ones are. Second, the

the sophistication of levels of representation that would be required in order to account for their manifested behaviours. See Povinelli (2000, p. 61-62) for discussion.

⁶⁴ Grice also notes that, for his proposed method to be of use to philosophical psychology, the genitor responsible for pirot construction had better collaborate with an *engineer* who can ensure that whatever psychological theories are being developed are of a design compatible with physics and physiology. “Since the genitor does not know about engineering, and since he does not want to produce futile designs, he had better keep a close eye on the actual world in order to stay within the bounds of the possible” (ibid., p. 141). This provides further evidence of Grice’s commitment to psychological reality.

laws being *ceteris paribus* laws allows for a progression of theories, desired to maintain continuity, as the more complex theories can contain the previous ones. We can imagine very general laws in the most basic theories, that gradually get more specific and rich in subsequent theories, with the more complex theories still being able to capture the effects of the general laws of the more basic theories.

Grice endorses the idea that “the psychological theory for a given type is an extension of, and includes, the psychological theory of its predecessor type” (ibid., p. 142). This idea poses a constraint on possible theories as we should be able to track the progression of particular concepts. That the laws of the theories are *ceteris paribus* laws allows for the psychological concepts gradually being fleshed out without contradicting previous predictions. The introduction of more complex concepts has to be justified by manifested behaviour. Note that as we are defining the theoretical concepts (WILLING, JUDGING, etc.) by their role in a theory, or in terms of the concepts of a theory, sequentially more complex psychological theories mean correspondingly more complex psychological concepts. This allows that both humans and other animals JUDGE in some sense; but preserves the intuition that different animals are capable of more or less sophisticated varieties of JUDGING. Grice indeed hopes to “safeguard the unity of psychological concepts in their application to animals and human beings” (ibid., p. 142).

It is perhaps worth presenting in some detail an example of the procedure for introducing psychological concepts (based on Grice 1975a, p. 134-137):

We observe a squirrel-like creature, Toby the squarrel, gobbling nuts in front of it. Given various other observations of Toby and his like we make generalisations such as *After a prolonged deprivation of nuts (and other substances) squarrels gobble nuts*. We decide such observations warrant psychological explanation and undertake to introduce an apparatus which provides an explanatory bridge between Toby’s having nuts in front of him and his gobbling them:

1. Given the relevant laws for a type of creature T are conjunctively L , we suppose that there is just one P , V , and J such that L . We label such instantiables P , 'prehending' (a kind of proto-perception), V , 'willing', and J , 'joining' (a kind of goal-satisfaction planning).
2. For a given T , there is a class N' ('necessities for T ') such that any N (in N') is a necessity for T just in case any member of T which suffers a sufficiently long non-intake of N will cease to exist.
3. If N is a necessity for T then *ceteris paribus* a moderately prolonged non-intake of N on the part of a member x of T will cause x to instantiate a particular sub-instantiable of V , V_i . x 's instantiation of V_i will be (*ceteris paribus*) quenched by a moderately long intake of N . (We can call V_i 'WILL for N ').
4. For any member x (e.g. Toby) of any type T (e.g. squarrels) there is a class of F' of object types ('objects familiar to x '); a class R' of relations ('relations familiar to x '); and a class A' of actions ('actions in x 's repertoire') which satisfy the following:
 - (a) that there are two ways, $w1$ and $w2$, (functions from F' and R' with respect to P) such that *ceteris paribus* if an instance of an object type (familiar to x) is related to x by a relation (familiar to x) then this causes x to instantiate a sub-instantiable of PREHENDING which corresponds in ways $w1$, $w2$ to F and R . We call the sub-instantiable of P 'PREHENDING F as R ' (e.g. 'PREHENDING nuts as in front').
 - (b) that *ceteris paribus* if, for some N (e.g. 'squarrel food') which is a necessity for x 's type, for some A in x 's repertoire and some F , R familiar to x , x has sufficiently frequently, when instantiating WILL for N , performed A upon instances of F which are R to x , and as a result ceased to instantiate WILL for N , then x instantiates a corresponding (N , A , F , R) sub-instantiable of JOINING we call 'JOINING N with A and F and R '.

Having introduced PREHENDING, WILLING and JOINING, a general 'PJV law' can be formulated such that N , A , F and R are not restricted to members of classes connected in certain ways with a particular creature or type of creature:

5. *Ceteris paribus*, if x WILLS N , PREHENDS F as R and JOINS A , F , R , with N then x performs A .

Grice introduces a notion of JUDGING (call it JUDGE_n) as derivative from JOINING: if x JOINS N with A , F and R then x JUDGE_n A , *upon-F-in-R*, *for* N (roughly: x JUDGE_n [acting in this way on that F gets N]). A simpler (not hypothetical) kind of JUDGING consequent on joining N with A , F and R , might be JUDGE_m: x JUDGE_m [F in R]. Using 1-5 it is possible to make an explanatory bridge between Toby's being in front of nuts and his eating them (assuming a sufficient deprivation of squarrel food).

4.2.4 *Towards mindreading*

We have seen how to introduce concepts for the psychological theory of a pirot like Toby the squarrel. We now need to consider how the theories get more complex to show how they accommodate more sophisticated behaviours. Ultimately we are interested in providing a theory that captures the behaviour exhibited by pirots corresponding to humans, in particular their mindreading behaviour.

The kind of step involved in pirot construction as we progress from one type of pirot to the next higher pirot is a process of 'content-internalization', whereby expressions that may appear legitimately outside the scope of psychological verbs come to be allowed within the scope of those verbs. The result of internalization in constructing the psychological theory for higher pirots is that it will be possible for the content of the psychological states of these pirots to include expressions like connectives, quantifiers, temporal modifiers, mood indicators, modal operators and, eventually, psychological states themselves.

There are two stages of internalization, as follows. Grice first notes that we can refer to psychological states in linguistic settings that have no direct relation with psychology, for example linguistic settings referring to temporal relations. So, a pirot can JUDGE [A], or could *in the future* JUDGE [A]. The JUDGING in both cases is the same operation. However, it is possible to incorporate the setting into the psychological state. Compare, *in the future* JUDGING [A] with FUTURE-JUDGING [A]. These are different kinds of judging. FUTURE-JUDGING is roughly equivalent to the idea of *expecting*. This is an example of first-stage internalization. A lower pirot might

have just an undiversified psychological state of JUDGING, while a higher pirot could have distinct PRESENT-JUDGING (equivalent to the simple JUDGING of the previous theory, and ensuring some continuity between the two theories) and also have PAST-JUDGING ('remembering') and FUTURE-JUDGING ('expecting'). Second-stage internalization involves replacing the intrinsic modifier which makes the particular psychological state of FUTURE-JUDGING [A] the state it is, with an operator within the content specification of a less specific state: JUDGE [*in the future* A].⁶⁵

Constructing successive theories for more advanced pirots will involve positing internalization where the explanatory bridge requires it. So, a creature which sits in front of two of its prey's potential hiding places monitoring their exits prepared to pounce in either direction, might be attributed with OR-JUDGING [prey in hole A, prey in hole B] via first-stage internalization from JUDGE [A] *or* JUDGE [B].

Second-stage internalization of the state of OR-JUDGING [A, B] would give rise to the state of JUDGING [A *or* B]. As Grice notes, OR-JUDGING [A, B] and JUDGING [A *or* B] are logically equivalent. So one could not posit second-stage internalization unless (or until) it is warranted by behaviour which requires the creature to combine operations, as in JUDGING [*if* A *or* B *then* C]. The presence of second-stage internalized theories is justified, for instance, by pirots requiring, to get by in their living condition, the ability for a high degree of inference where conclusions are derived from premises "within a single thought": that is, when for the "explanation of the ranges of behaviour to be assigned to that type of pirot, the pirot has to be capable, for example, of thinking [since, A, B, and C, then D]" (ibid., p. 149).

⁶⁵ There is a certain looseness in Grice's description of the internalization process which is acceptable at this point in the proceedings but of which we should be aware. Note that saying '*in the future*, x JUDGES A' involves the deployment of our own (English speaker's) concepts of temporal reference. However, temporal concepts may well not be universal, so when we posit internalization of temporal concepts for a creature, strictly speaking we should be talking about pirots being able to JUDGE [F' A], where F' would be some kind of futurate operator which would need to be specified for the particular creature based on its behaviour.

Second-stage internalization, then, in some cases at least, entails hypothetical and temporal inferential abilities. It can also entail metarepresentational abilities, allowing creatures to make inferences explicitly. Having issued a caveat about not increasing complexity of theories beyond necessity,⁶⁶ Grice goes on to consider the possibility of the internalization of psychological states themselves. We can assume that much the same rules apply as above: a creature is credited with first-stage internalization or second-stage internalization of psychological concepts as warranted by manifested behaviours. Grice largely passes over the details of the first phase; he is particularly keen to focus on second-stage internalization of psychological concepts as this is what he sees as bearing on the philosophical problems he in part set out to address. We will return to first-stage internalization of psychological concepts as it is relevant to our concerns; first let us recognise Grice's interests to further illustrate his system.

In these later stages of the programme in *MPP*, we encounter highly rational creatures. Thus Grice introduces JUDGING*, a kind of JUDGING defined in terms of WILLING. Informally, " x JUDGES* p just in case x WILLS that, if he has to choose between a kind of action which will realize some end of his just in case p is true and a kind of action which will realise that end just in case p is false, he should WILL to adopt some action of the first kind" (ibid., p. 153). Grice doubts that JUDGING* is a substitute for JUDGING; it would seem that Grice intended that we take JUDGE* just as another form of JUDGING (a kind of 'all things considered' belief state) which specifies a state undifferentiated in lower pirots:

I doubt if one would wish judging* to replace the previously distinguished notion of judging; I suspect one would wish them to coexist; one would want one's relatively advanced pirots to be capable not only of the highly rational state of judging* but

⁶⁶ "Principles of parsimony dictate that we introduce no explanatory tool more powerful than is required to do the job; when psychological states which have been generated by first-stage internalization can be seen to be adequate for explanatory purposes, we should not attribute to a pirot the states which would be reached by superimposing second-stage internalization" (ibid., p. 150).

also of the kind of judging exemplified by lower types, if only in order to attribute to such advanced pirots implicit or unconscious judgments.

(ibid., p. 152)

On the questions of privileged access and corrigibility, Grice motivates the evolution of creatures who are at an advantage if they can JUDGE that they JUDGE (or WILL) something without having to observe behaviour. This brings about the advantage of self-control, but only if the JUDGINGS about JUDGINGS and WILLINGS are accurate. So creatures which evolve in this way would have to be such that, other things equal, if they JUDGE (WILL) [A] then they JUDGE that they JUDGE (WILL) [A]. And this is just a form of privileged access. In order for such a self-controlled creatures to be successful creatures, Grice continues, the evolution would have to be such that (other things equal) if they JUDGE that they JUDGE (WILL) [A] then they JUDGE (WILL) [A]. The suggestion is that if we then define 'x believes that p' as 'x JUDGES that x JUDGES that p', we get the result that x believes that p iff x believes that x believes that p.⁶⁷

So, central to this part of MPP is a discussion of what kind of theory would underwrite a respectable implicit definition of belief in a way that accommodates philosophical intuitions and precepts. The kind of theory is one that assumes that behaviour is prehended in certain complex ways. Let us return to first-stage internalization to illustrate. Suppose there are creatures which display behaviour which we could describe as plan- or intention-thwarting behaviour (a she-cat seeing off a predator stalking her cubs; a dominant ape intercepting a would-be suitor of one of his females). One way of explaining this behaviour would be to suppose that while the interloper (the predator or the suitor) JOINS some need (food; mate) with an action type (stalking; courting) and a configuration of individuals (cubs; female apes), the operant displaying thwarting-behaviour (the she-cat or dominant ape) may be

⁶⁷ Although it is not explicit in the text, I assume that Grice's defining 'x believes p' as 'x JUDGING that x JUDGES that p' implies that there may be as many forms of belief as there are forms of JUDGING. Grice is however particularly interested in JUDGE*, and other highly rational concepts. See Grice (2001, p. 77-80).

JOINING a need with an action-type and a configuration of individuals such that it represents the interloper JOINING needs with actions and individuals-in-play. In such a case, we would have the operant JUDGE_n-JUDGE_m-ing (roughly: JUDGING that the interloper takes the cub/female ape for its N). Of course, in either of these cases, the same thwarting behaviour could no doubt be explained in more parsimonious terms.⁶⁸ Nevertheless, note how we could have quite a respectable attribution of psychological concepts to a creature without having to ask whether the creature has a concept of *belief*.

However, there is a problem with this outline of a process of first-stage internalization of psychological concepts. To suppose that a creature could be JOINING together object-types which include *another individual JOINING something* involves a creature PREHENDING another creature performing an action which realises this plan. In theories of earlier pirots, we could assume that the relevant action descriptions be cast in non-theory laden terms, just as patterns of movements, etc. The fact that the same pattern of movements (a man raising his arm) could be a manifestation of different actions (a man waving away a fly and a man trying to attract the waiter's attention) could be overcome perhaps by reference to the fact that a JOINING involves a behaviour (understood just as a pattern of movement) and a need (or end). But a creature cannot PREHEND another creature JOINING a need with *an action on F-in-R* just on this basis. Joinings are not observable. For a creature to be JOINING joinings, it needs to be able to discriminate patterns of movement *as* actions, i.e. in a theory-laden way.

⁶⁸ For any attempt to implement the programme in *MPP* one important task would be to devise experiments that would allow the theorist to discriminate between possible psychological theories. On *seeing* and *knowing* in chimps, see e.g. Povinelli & Eddy (1996); Povinelli, Bierschwale & Cech (1996); Tomasello, Hare, & Agnetta (1999); Hare, Call, Agnetta & Tomasello (2000); Hare, Call, Tomasello (2001); Tomasello (2002).

Grice is of course aware of this feature of the programme in *MPP*.⁶⁹ However, he does not provide any reason for it, beyond the indirect support the assumption receives from the fact that the resulting theory is successful in underwriting the philosophical notions, as desired. While this may be sufficient for Grice, an ethnoscientific investigation into commonsense psychology cannot appeal to this type of support. Providing an analysis of *belief* is not what motivates our concerns. Our aim is to investigate I-psychology and the concepts employed therein and, we have argued, there is no reason to suppose that the vernacular-commonsense notion of 'belief' will be part of that account. Rather, it will be suggested below that vernacular-commonsense ideas about the mind result from interpretations by more general conceptual systems of the 'structural descriptions' licensed by I-psychology. That is, while the features of I-psychological description are hard-wired and universal, reflective commonsense interpretation of these (as per the many senses of the vernacular terms 'believes', 'desires'; 'thinks', 'wants') will be culturally and individually variable.

To sum up: while the aims of *MPP* diverge from ours, the paper is rich in suggestion about how to proceed in constructing a naturalistic account of commonsense psychology. Rather than just attempting to systematize everyday statements and intuitions about psychological matters into some kind of underlying laws, we devise a framework according to which agents are attributed with psychological concepts where necessitated by the activities they engage in. A creature, call it 'John', evinces certain forms of behaviour (mostly when interacting with others of its type) which necessitates that we posit, say, that its plans bear on the plans of others. We say of John (and his type) that he has internalized certain psychological concepts – those

⁶⁹ Indeed, it is a central plank in his argument against eliminativist tendencies towards the end of the paper (around *ibid.*, p. 159), where Grice argues that precisely because the same set of movements can be realizations of different actions, psychological explanation must involve agents perceiving actions as such.

that are necessary for a bridging explanation appropriate to his behaviour. We do not say of John that such activities are possible because he has (or knows) a theory of behaviour in terms of vernacular-commonsense notions such as belief and desire.

We accept, along with Grice, that an account of this kind of John's activities will presuppose conceptualization of behaviour not merely in terms of superficial descriptions but in terms which individuate actions as such (i.e. in terms of plans, goals, etc.). In the next section I outline how the ethnoscience of commonsense psychology might make use of Grice's programme.

4.3 AN OUTLINE

We have seen from the discussion of Grice's proposals that a psychological theory for a creature that has internalized psychological concepts presupposes that such a creature can apprehend behaviour in theory-laden terms. This means that rather than just supposing that the creature perceives behaviour simply in terms of patterns of movement, creatures with the relevant cognitive abilities must perceive behaviour under a far richer description. It seems we would have difficulty not presupposing this because the kind of laws in terms of which we would need to define states such as JUDGE_n-JUDGE_m-ing would not otherwise operate at the level at which they appear to operate in humans. To adapt an example from Grice (1975a, p. 159), for a (socio-culturally advanced) creature's pre-emptive strike behaviour to be successful, the creature must be able to discriminate between cases where another creature is advancing onto it and those where another creature is merely advancing towards it.

It may seem, however, that to suppose that our advanced, psychologically sophisticated creatures already perceive behaviour in this way is just to suppose much of what it is we are meant to be explaining. While this is more or less what we shall be doing, there are two main points that should be borne in mind as my proposal unfolds. First, that the problem about discerning behaviour in theory-laden terms is solved by supposing that I-psychology is an internalized theory of the sort that Grice

envisages. Second, that vernacular-commonsense terms do not come as part of I-psychology but as interpretations of what I-psychology does.

We will assume that the kind of psychological laws that we posit for creatures that can engage in human-like social interactions are ‘cognized’ by human-like creatures, and that their joinings (or plannings) are made according to these laws.⁷⁰ This will mean that joinings (or plannings) involving the joinings (plannings) of others will be underwritten by an understanding of others in terms of their own theory for action.

Given the complexity of the task that faces us in describing such a theory we would have to assume that much of this knowledge is tacit. It does seem that humans have partial access to aspects of the theory and are able to talk about some of that. But to take advantage of the benefits of being able to talk about mindreading matters (reporting psychological events, for example) it would not be necessary to talk about all of the theory – talking about mindreading at a more general level would probably suffice. And, it appears, *how* people talk about mindreading is not particularly significant for the content of the theory; for example, many languages have no word for belief and lack a suitable alternative, but speakers of these languages do not therefore fail to mindread (Rhum 1993). Cross-linguistic variation of mentalistic talk may well contrast with cross-cultural stability of the final state of (some part of) the psychology faculty.

Let us consider the implications of this stance in 4.3.1, before returning to fill in some of the details of the picture in 4.3.2. In 4.4 I begin to anchor my proposal in current research.

⁷⁰ This idea echoes Fodor’s claim that folk psychology is true and innate, in his words that “*homo sapiens* is [...] uniquely the species that is born knowing its own mind.” (Fodor 1989, p. 133).

4.3.1 *Methodological preliminaries*

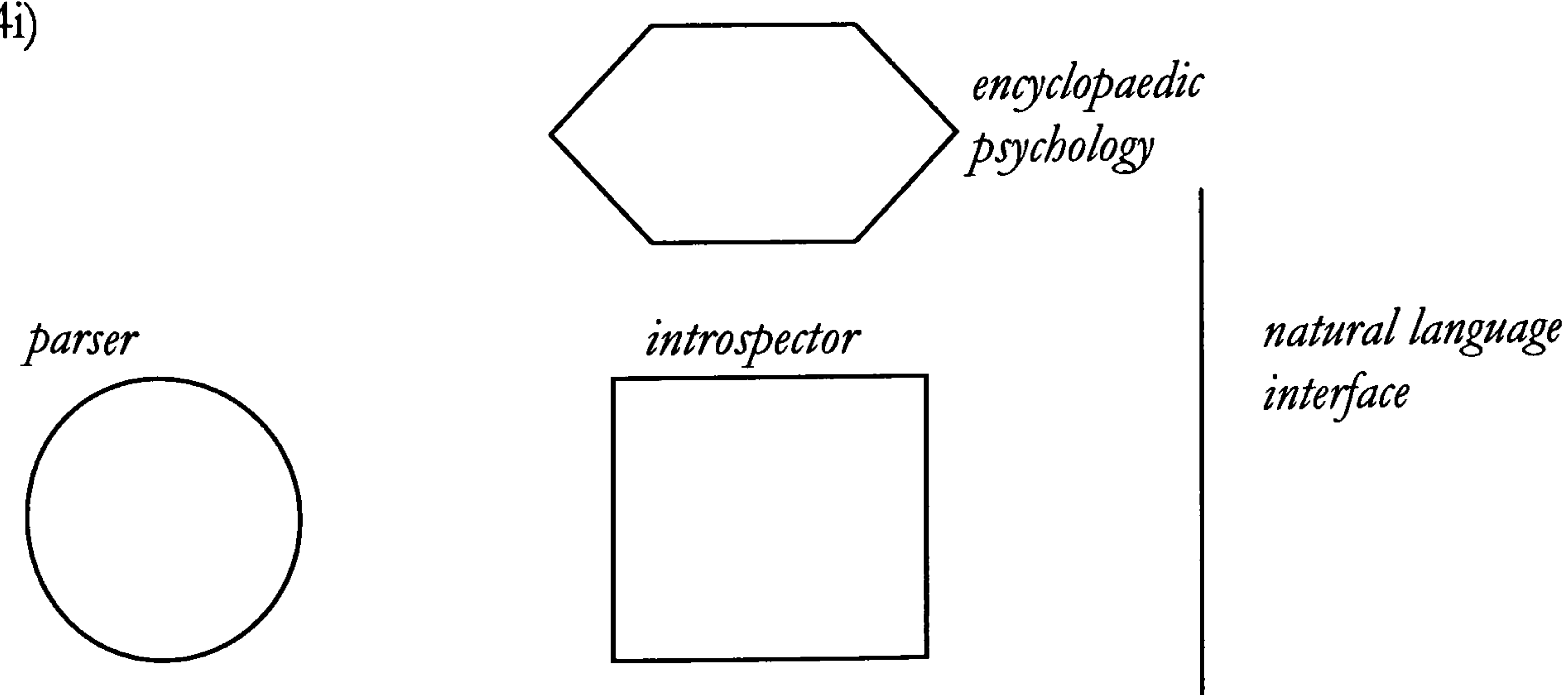
Other components [of the mind/brain] provide ‘common sense understanding’ of the world and our place in it – what is often called ‘folk psychology’ and the like, though we should be careful to observe the practice of serious ethnoscience, distinguishing parochial and culture-bound notions from the elements of ‘folk theories’ that are a common human endowment, ‘a direct gift of Nature’; not an easy problem, and one that is, I think, too lightly dismissed.

(Chomsky 1993, p. 34)

We are exploring the framework that is used on a day-to-day basis to understand behaviour. But we have acknowledged that not everything that is called upon to do this is likely to have a unified explanation. Let us try and portray what components might be involved in delivering the kinds of abilities that mindreading covers. In the next sub-section I consider in more detail specific aspects of I-psychology and how the psychology faculty interfaces with other systems.

For the sake of argument I will make a concrete suggestion that is in keeping with what is known and highlights my main concerns. Figure (4i) provides an initial illustration. As a starting point for discussion, I will proceed by describing the individual parts of the picture.

Fig. (4i)



The *parser* implements knowledge that is largely innately determined and universally follows identifiable developmental stages in normal individuals. What this information is, is a matter for empirical discovery, and may involve complex rules or algorithms. The knowledge on the basis of which the parser operates is likely to feature concepts that have only limited counterparts at an accessible level. The I-theory for this ‘behaviour parser’ will be in whatever form the best theory dictates; the kind of information suggested by Grice (1975a) introduced above is one possibility. We can think of the behaviour parser as implementing I-psychology. As we saw for language in 2.4.1, a full account of the psychology faculty would perhaps need to incorporate interface issues. I explore these ideas further in 4.3.2 immediately below and discuss experiments by Baldwin (e.g. Baldwin et al. 2001) on how children parse behaviour streams into action segments in 4.4.

The *introspector* allows for reflection on the results of the operations of the parser and its associated systems, in terms that we (as individuals rather than theorists) can reason about. This is the level at which vernacular-commonsense psychological concepts are in operation. The concepts deployed by the introspector will depend in part on whatever knowledge there is in the parser, and in part on the reflective resources and circumstances of the individual. The relation between the I-psychological system and the introspector may in part follow some developmental pattern and in part be moulded by environmental factors, including perhaps amount of practice, cultural constraints, and natural language (see 5.4). These factors are significant because some of the motivation for the introspective level comes from the utility for individuals not only of reflecting on and storing information about psychological matters, but also of communicating this kind of information.

The inferences of the introspector are presumably represented in the medium of a conceptual representation system or language of thought (Fodor 1975). The next step is to consider the *interface between the introspector and natural language*. Here there is a

general question about the relation between natural language terms and the concepts they can express. Sperber & Wilson (1998) point out that there is unlikely to be a one-to-one correspondence between concepts and lexical items. We might expect their proposals about the mapping between the mental and the public lexicon and the formation of *ad hoc* concepts to apply to the resources that we use to reflect on psychology just as much as it holds of other cases. For example, different languages do not have the same range of words and expressions to pick out mental states, but speakers of those languages are still equipped to understand roughly the same situations. In English a variety of words and expressions ('think', 'know', 'believe', 'hold'; 'want', 'desire', 'hope for', 'aspire to') could correspond to fewer quite similar mental states; conversely, at the introspective level concepts might be in operation for which English lacks a word. Also, in cases where English has a word for some mental state, the same word might stand for different concepts on different occasions⁷¹. In any case, it is clear that natural language is not necessarily a reliable indicator of the concepts that are operating at the level of the introspector, let alone of the knowledge represented in the behaviour parser.

A further body of knowledge related to psychology, which is almost certainly called upon in making more complex judgements about behaviour (although whether accessing it improves the success of resulting attributions is probably a moot point), we might call *encyclopaedic psychology*. Encyclopaedic psychology is information that is related to understanding behaviour and can be used to do so – it contains, for instance, useful generalisations, scripts of typical scenarios, stereotypes of attitudes and personality traits. It may also include knowledge picked up from popular science

⁷¹ Some of the concepts may be incomplete or deferential, or the result of cultural elaboration; some may be idiosyncratic concepts formed on the hoof in the process of a particular interpretation, and subsequently become stable as opportunities for their re-application occur (see Carston 2002, ch. 5).

books and programmes and news reports; some version of information learnt academically could also be expected to feature here.⁷²

Encyclopaedic psychology interacts with the introspector at least in the first instance when the information is stored, as information gets stored *as* encyclopaedic psychology according to its relevance to the psychological. It is thanks to the concepts of the introspector that the information can be recognised as psychological. The information in encyclopaedic psychology may even be organised in a pseudo-theoretic way; if so, the introspector could play a role in such systematisation, imposing a structure on it so that it is in the most readily usable format for when it is called upon for various uses. There could be formatting and organisational issues that are innately determined and specific to psychology.

Clearly, however, knowledge in encyclopaedic psychology does not have the universal character of the knowledge in the parser. Nor is it similar to what is inside the introspector, which, though subject to more individual and cultural variation than the behaviour parser, is still closely constrained. Not everybody builds an encyclopaedic psychology, and there must be huge variation in what it contains and how good people are at applying it. Here we have an example of something that can be called upon in predicting and explaining behaviour but that should be kept separate from other types of psychological knowledge. It is acquired differently and may be put to use in a different way (optionally, consciously, in a problem-solving kind of way rather than a fast and automatic way).

For various phenomena related to commonsense psychology (as conceived of pre-scientifically), we can ask which of these elements they are primarily due to, or investigate how the elements interact to produce the phenomena. Different theorists have been interested in what are effectively different aspects of so-called folk psychological understanding, or mindreading abilities. For example, many philosophical accounts of 'folk psychology' relate to the interface between the

⁷² *Encyclopaedic psychology* should not be confused with general encyclopaedic memory.

introspector and natural language, informed by insight from encyclopaedic psychology. Developmental psychologists tend to be more interested in the ‘parsing’ end of the spectrum. Unsurprisingly, appeals to tacit knowledge and modularity are more convincing in the latter case than the former. This underlines the point that ‘mindreading’ is an umbrella term for all kinds of mental processes called upon in psychological understanding, and that we should be clear about which of these we intend to refer to in discussion of commonsense psychology. How much of what is represented in figure (4i) should be construed as part of a Psychology Faculty is an open question. I turn now to concentrate on the study of the aspects of mindreading in which I am most interested.

4.3.2 U Ψ , I-psychology and the Psychology Faculty

I want to suggest that there is an equivalent to UG for commonsense psychology. U Ψ , the core component of the Psychology Faculty (FP) is, roughly speaking, an internal theory of acting. The I-component of the core of FP will contain specific laws for plans (and their related needs) – as many as there are in an individual’s repertoire. In some cases, the I-laws relate to plans which bear on the plans of others; these laws will specify such features in the same terms as for the self. This gives rise to the type of recursiveness which is a hallmark of commonsense psychological thinking. Spelling out such a proposal in detail would be an undertaking which would take us well beyond the scope of this thesis. What I am hoping to achieve here is to clear the ground so that a programme of this sort can be entertained, provide reasons for accepting the basic idea, point towards existing work which is compatible with it, and explore some of its consequences. However, I will consider some possible directions in which a programme of this sort might be pursued. Of necessity, the discussion will be schematic.

The I-theory that arises out of U Ψ is part of the cognitive system. As with language, we can think about this I-theory on the analogy with a computer program embedded in a larger system, interfaced so as to enable the system to perform its

function; in this case the execution of plans in perceived circumstances. Given that the perceived circumstances can involve other agents engaged in actions/activities, there will be both a 'production' and a 'comprehension' component of the system. Performance systems for production will involve executing motor-schemas in relation to awareness of current circumstances; in this regard performance clearly will involve other components of the cognitive system (spatial cognition, etc.). Comprehension will also involve the co-ordination of motor-schemas and awareness of current circumstances. In some ways analogous to the case of language, the I-theory of (this part of) FP can be thought of as providing pairs of structural descriptions: IF (for internal, or first-person form) and EF (for external, third-person form). IF is a description of the first-person dimension of actions, motor-schemas and perhaps sensory feedback and other aspects of phenomenal experience plus 'first-person' representations of scenes. IFs would be available for all creatures, as they are necessary to co-ordinate the execution of movement with objects in the world. EF is a description of the external aspects of action, including situation descriptions of stereotypical beginning and end-states and descriptions of patterns of movement. Where a scene contains another agent acting, the parse for that will result in an EF-IF pair. The fact that action is conceived of as a pair of internal and external representations which crucially are linked may explain how imitative learning arises.⁷³

The relation of these 'structural descriptions' to the level of the introspector, and wider consequences for the nature of the psychology faculty, will be discussed in 4.5. Let us first begin to make this proposal more concrete by looking at evidence about how humans segment the streams of behaviour they perceive.

⁷³ For discussion of evidence from word-learning about the development of action concepts in 12 to 18-month-olds which is consistent with this picture, see Tomasello (1995, 2000). For discussion of the difference between apes and children with regard to action concepts and imitative learning, see Tomasello 1999.

4.4 ACTION PARSING

When *parsing* is applied to language it generally means the process by which a linguistic stimulus is segmented and analysed (or assigned a structural description) according to the principles of an I-language (see Dowty, Karttunen & Zwicky 1985). These principles superimpose a structure corresponding to word, phrase and clause boundaries onto a superficially undifferentiated stream of sound. Although I am concerned mainly with speech, the stimulus could be graphemic or visual as well as acoustic. In fact, it may be that the language faculty applies itself to absolutely anything, or at least attempts to do so. It will however only generate anything interesting when the stimulus it applies to is in fact linguistic (i.e. determined by a similar system), as opposed to just background noise (e.g. a passing taxi, a squeaking door). The process is fast, automatic and unreflective. It is this structure that is assigned an interpretation in comprehension.

In appealing to the notion of parsing with regard to understanding behaviour, I have something similar in mind. In a general sense, it is informative to talk about understanding behaviour in terms of action parsing because the terminology suggests a commitment to the idea of a ‘theory’ (I-psychology) applied automatically to social situations which provides initial analyses of behaviour quickly and effectively. It may be that, like I-language, the theory is applied in more situations than these, but it will give the best results when it is put to work on some input it recognises. An undifferentiated stream of behaviour thus gets segmented into units which correspond to actions.

For this idea to work, there would have to be some correlation between salient features of observed behaviour patterns and underlying intentions, and people would have to latch onto these features rather than others in deciding what counts as an action boundary. As it turns out, 10-month-old infants already appear to segment streams of behaviour, and do so according to perceptual clues, such as changes in

direction of movements and eye-gaze, rather than by appealing to, e.g., time slices of action (Baldwin et al. 2001). And these preferred cues do indeed match up with intentional structure in a way that, say, ten second segments of behaviour would not.

Before considering the experimental evidence for these claims, it is worth pointing out that there are more specific ways in which the parsing parallel is insightful. Just as you don't need to know what words mean in order to recognise boundaries between lexical items, perhaps children don't need to have fully-fledged concepts of the actions they observe in order to recognise action boundaries. A comparison with phonology will illustrate this point. Furthermore, parsed 'behaviour strings' may need to be interpreted and further integrated with contextual information before complete understanding is reached, just as the structural descriptions given by the linguistic parser need to be interpreted using semantic and pragmatic information.

Let us concentrate on action parsing to start with. Support for the hypothesis that humans discern structure in behaviour comes from recent work done by Baldwin and her colleagues. Having familiarised small children with a video clip of an actor engaged in a straightforward kitchen activity (picking up a tea-towel from the floor and hanging it on a peg; putting away in the freezer a tub of ice-cream from the counter), the children were then shown different versions of the same clips – *completing* test videos where the action was freeze-framed at points coinciding with completion of one intention and the beginning of another; and *interrupting* test videos where the scene was frozen mid-intention. The children looked longer at the interrupting test videos, suggesting that they have expectations about intention boundaries that are confounded by the mid-intention pauses. The format of the experiment is similar to language processing studies where subjects are presented with linguistic strings with beeps inserted at different points (Fodor & Bever 1965). In these cases, subjects tend to displace beeps towards phrase boundaries, indicating that their preference for segmenting strings in this way overrides other perceptual interference.

Having established that there is sensitivity to intention boundaries, Baldwin goes on to note that this experiment does not show that infants necessarily are able to see the intentions behind the actions, but it does show that they already identify units that will be useful in further processing. The suggestion is that children may register completed actions on the basis of kinetic regularities,⁷⁴ and then be able to superimpose intentional structure to understand what is going on in a particular instance once the appropriate concepts are in place.⁷⁵ In this respect the action parsing is going on at a level comparable to phonological analysis in the case of language processing. It is recognised that babies are sensitive to sound patterns associated with different languages well before they have begun to process words and their corresponding meanings (see Mehler & Dupoux 1994). Similarly for behaviours, infants may be sensitive to the statistical regularities of movements before associating such movements with relevant underlying causes.

In the action study, choices about intention boundaries for the completing test videos was agreed by a number of adults. Indeed, according to Baird & Baldwin (2001, p. 194), research into adults' perception of intention boundaries appears to show striking agreement between different individuals over what count as units of action (although there is less agreement between theorists as to how these units are individuated). As mentioned in 4.2 and discussed in detail in 4.5.1 below, there is more than one way of describing the same observed behaviour. Baldwin et. al (2001) also make this point, noting that in the test videos the actor could be described as having a general intention to tidy the kitchen, as specifically wanting to pick up the towel, or on the basis of the individual segment-goals of bending over, grasping the towel, standing, hanging it up. A reported study undertaken by Baird et al. (2000) indicates that adults easily switch between these levels. Young children, at least initially, may have less room for manoeuvre. Without the concepts of actions and intentions that the parsed behaviours are to be mapped onto they might only focus

⁷⁴ See Wynn (1996); Sharon & Wynn (1996)

⁷⁵ This strategy would enable at least partial processing even in the context of novel scenarios.

on the movements of the component actions. However, one could also imagine it sometimes working the other way: if the overarching intention is quite basic but being performed in some novel way, the recognition of the basic overarching intention could inform the individuation of the component actions.

It would be interesting to explore whether there is adult agreement not only in what can count as a boundary at these different levels, but also about which is the significant level to be considered in understanding some piece of behaviour. Relevance theory (Sperber & Wilson 1986/95) predicts that we tend to construct the most relevant representations of stimuli that are initially picked out as worth processing (although of course, what may be the most relevant representation to one individual may not be the most relevant representation to others). So we might expect that agreement over what the significant level to be considered actually is, should depend on the particular activity in which the observed agent is engaged. As it turns out, this prediction is confirmed by further studies on action parsing undertaken by Newtonson and colleagues (on whose initial results Baldwin's experiments were based). Having established (Newtonson 1973) that people reliably segment behaviour according to instruction (subjects consistently picked out "the smallest actions that seem natural and meaningful" and the "largest actions that seem natural and meaningful" at five week intervals, and the fine-grained units of analysis turn out to be sub-parts of the larger units seen as super-ordinate goals), Newtonson (1976) reports that:

each individual has some 'range of analysis' within which he may operate. That is, each person can analyze a sequence of behavior into fine units, large units, or at least one level between the two. We have consistently found that, when instructed to analyze a sequence 'naturally', the mean number of units for a given sequence will be somewhere in-between that for fine-unit or gross-unit instructed subjects. Where on that continuum the level of analysis falls is very much a function of the particular sequence. In general, natural-unit analysis for sequences portraying highly organized, step-by-step action, with a clear hierarchy of subordinate and superordinate goals, will tend to be closer to gross-unit levels. Irregular, loosely

organized action sequences will tend to produce natural unit sizes closer to fine-unit analysis.

(Newtonson 1976, p. 231)

It might be illuminating to reconsider quite a lot of this and similar work of various schools of 'social psychology' from the perspective of a relevance-theoretic framework, with a view to assessing contributions to mindreading abilities made by the processes this research taps, although I do not intend to do so here.

To return to our present concerns however, as noted above, perhaps we can pursue the parsing analogy further. Specifically, in the case of language, once the parser has assigned a structural description to some utterance, what you then do with your analysed string depends on further interpretive processes. Both determining the proposition expressed and recovering any implicit content relies on the application of communicative principles to the output of the parser to work out the speaker's intended meaning. Perhaps we are in a similar situation with our initial analyses of behaviour. Having segmented the behaviour stream into action units, we are then left with the task of working out the corresponding intentions. We might think of this as a kind of relevance-driven 'inferential' stage to commonsense psychological understanding.⁷⁶

So, for any specific instance of commonsense psychological understanding which might be put forward for theoretical analysis, we should bear in mind that, as with language, at least parsing and interpreting and contextual evaluation may have contributed to the final decision about what some observed behaviour pattern 'meant'. Of course, understanding a whole situation does not always just involve parsing individual behaviours. Often we have to consider not just the intentions of one person but the intentions of many interacting agents in a scenario. To do this we will both have to work out the relevant plans and goals of each person and compute

⁷⁶ Note that this is probably an unreflective process, comparable to assignment of reference to a pronoun.

how this information compares with alternative representations of how the world is and alternative expectations about how the world should be. We are very good at solving these kinds of ‘social equations’ too, as well as those where there is just the behaviour of an individual to take into account. As well as the ability to parse actions, the ability to parse *interactions* also deserves attention.

This involves being able to work out what follows from combinations of embedded attributed mental states (joinings/plans). I discuss how we ‘mesh’ different attributed representations of high degrees of complexity very fast and semi-automatically according to almost logical patterns of inference in more detail at the end of chapter 5, where I consider both communicative and non-communicative examples to illustrate the point. For now, note that arguably the very presence of two individuals in a situation already involves dealing with an ‘interaction’ of sorts, even if the two are not strictly speaking *interacting*. The presence of an observer to some behaviour, even if the observer is not a participant in the behaviour, already alters the dynamic of a situation; contrast the behaviour of an actor being observed unwittingly. Watching an individual behave when you are not part of the scene is different from watching behaviour when the person who is being watched knows they are being watched. Once there are two people involved in a social environment, even if they are not directly interacting, the fact that each recognizes that they are not alone justifies some level of interpretation of their behaviour, even if it does not create or justify an expectation of optimal relevance as ostension does.⁷⁷

⁷⁷ See Wilson & Sperber (2002) and references therein for details of the relevance-theoretic comprehension procedure. Briefly, according to relevance theory, human cognition is *relevance* oriented (that is, we pay attention to stimuli that give rise to greater cognitive effects for less processing effort), and utterances, ostensive stimuli which demand attention, create expectations of *optimal* relevance, thus justifying a comprehension procedure that excludes all but at most one interpretation. Upon recognition of ostensive behaviour we process a given stimulus according to a particular strategy: construct an interpretation following a path of least effort, considering alternatives in order of accessibility, and stop when there are enough effects to satisfy the expectation of *relevance* that the behaviour conveyed. See also Sperber (1994b).

In the next section I draw some distinctions about possible levels of understanding of behaviour which will be useful as we proceed.

4.5 UNDERSTANDING BEHAVIOUR

If we think about the explanations we can provide of the behaviour we observe, it is possible to discern at least two sorts of judgement that feel rather different. Some judgements are snap-judgements – of the ‘what-will-happen-next’ variety. Some are more considered.⁷⁸ The same situation can give rise to both kinds of judgements.

Stich & Ravenscroft make reference to something that sounds like the kind of thing I have in mind for snap-judgement in the list of ‘folk psychological’ abilities that they take to require an explanation. Together with describing people in psychological terms, providing explanations, making predictions and recognising folk psychological generalizations, they mention “the capacity to *anticipate* other people’s behaviour without saying anything at all” (Stich & Ravenscroft 1996, p. 125) and offer as an example “watching a baseball game, our eyes immediately jump to the hot-tempered manager when the umpire throws his star player out of the game” (ibid.).⁷⁹

While we probably could spell out the dynamics of this situation if someone asked us to, we don’t actually have to provide an explicit characterisation of what is going on in an example like this in order to understand the observed behaviour enough to both make sense of it and have expectations about its consequences.

⁷⁸ Perhaps snap-judgements are linked directly to an emotional response, or predictor of emotional response, while the others provide a worked-out explanation for why the response predicted or experienced was the one it was.

⁷⁹ I would like to leave the issue of linguistic behaviour to one side for the moment. I think that despite the fact that the other four kinds of folk psychological practice listed are most often linked to verbal reports, the reference to “not saying anything at all” is a red herring. As we will see, what makes the anticipation cases different for me is that they may not go up to the level of the introspector at all, not whether, having got there, they get as far as the interface with natural language.

Indeed, in some cases it can be quite hard to articulate all the details of the mental state ascriptions and other evidence that you call upon to evaluate a situation. Sometimes doing so is insightful, sometimes it can take more effort than it is worth. Certainly, some people are better than others at providing the explicit characterization of an observed situation, whether or not it is useful to do so in that instance.

4.5.1 *Levels of analysis and levels of understanding*

Given just the physical manifestation of some piece of behaviour we can usually think of more than one analysis of it. There is a straightforward way in which this observation holds whereby the same piece of behaviour can have completely different causes on different occasions. So, if someone reaches out their arm they could be trying to catch the waiter's attention, or just waving away an annoying fly. These two explanations are unlikely to be correct at the same time. There is however another possibility where there may be two ways of describing some piece of behaviour which *are* compatible. I don't mean cases where there is more than one goal behind a piece of behaviour (so not cases where the fly and the waiter are conveniently repelled and attracted with a single gesture). What I have in mind are cases where the same goal can be described (correctly) in more than one way. So if we ask why the detective ducked behind the parked convertible the answer might be (a) *to get out of the cop's line of sight*, or something more like (b) *because he didn't want the cop to see him*, or even (c) *because he didn't want the cop to know he was tailing him*.

Note that these cases are not quite the same as cases where there is an overarching intention with component parts. For instance, an intention to get a beer from the fridge involves going to the fridge, grasping the doorhandle, opening the door, selecting a beer, closing the door; and an observer might focus on the action sequence as a whole individuated by the desire for beer, or on the separate sub-actions individuated, say, by sub-goals or changes in direction of movement. Focusing on some particular point of the action sequence, it might well be correct to say both that the drinker is opening the fridge and that the drinker is getting a beer.

Of course, the detective example can also be broken down like the beer example: on the one hand, the detective might have an overarching intention to solve the case, and not being seen could be crucial to achieving it; and on the other hand ducking behind the car might involve sub-components (lengthening stride, squeezing past parked truck, crouching behind wheel). But here I am interested in a slightly different point again. The unit of action that we are concentrating on is the action construed as ducking-behind-the-car. And it is that individual unit that has different (compatible) readings. Crucially, not all of these readings need involve psychological concepts. Those readings that do involve psychological concepts can be more or less complex. Option (a) above makes no reference to psychological states whereas (b) does; (c) further involves embedded psychological states.

So, we have distinguished three ways we could consider some piece of behaviour to have different descriptions. The same observed behaviour could be triggered by different underlying causes (the fly or the waiter); the same observed behaviour could be perceived either as an individual action or as a sub-component of a more complex action (the beer case); the same observed behaviour could be described at different psychological levels (the detective case). It is these latter cases that concern me here. A very old joke in (1) provides another example:

- (1) Q: “Why did the chicken cross the road?”
 A: “To get to the other side.”

Here, the answer to the question undercuts any explicit psychological explanation we might be in the process of constructing, giving rise to some (comic) effect; at least, children usually amuse each other with it.⁸⁰ Another case is the stereotypical bad actor in (2) asking :

- (2) “What’s my motivation?”

⁸⁰ It is likely that children begin to find this funny at around the same time they begin to reflect on psychological states as opposed to just recognising and reacting to them.

when the scene he's shooting just involves some banal self-contained task like brushing his teeth or drinking a glass of water. In a sense these examples are a pair: in the first one the observer is attempting to 'read' more into the situation than was present, and in the second the agent is attempting to 'write' more into the performance than is likely to be recovered.⁸¹

4.5.2 *Complex and basic behaviours and interpretations*

All animals need to form some kind of representation of the world in order to interact with it (more or less successfully depending on the quality and sophistication of their representations and the complexity of whatever tasks they are called upon to perform for their continued existence).⁸² We share some aspects of cognition with other creatures. We have comparable if not similar visual systems; probably we have similar reactions (to some extent) to things in our cognitive environment, just as we can have similar reactions to things in our physical environment. The attention of other animals is geared towards noticing bright lights and sudden loud noises much as ours is, and probably for the same reasons.

Of course, similar reactions do not guarantee similar underlying causes.⁸³ The cognitive resources of humans with regard to mental representations appear to extend further than those of other animals. Humans solve complex problems which

⁸¹ Using these examples we can say something about what might govern the search for hypotheses about interpretations of behaviour. For instance, in neutral contexts you get enough contextual effects for the effort expended in processing the stimulus from assuming that drinking a glass of water is motivated just by thirst, so anything beyond that will be lost (unless something striking about the way in which the water is drunk catches the observer's attention and leads to a search for some further nuance in interpretation). Conversely, *getting to the other side* practically defines what *crossing the road* amounts to, further (psychological) contextual effects are required to satisfy expectations.

⁸² See Gallistel (1989, 1990); Hauser (1996, 2002); Bekoff & Jamieson (1996); Allen & Bekoff (1997) for discussion of the cognitive capacities of various creatures.

⁸³ See section 4.5.4 below for discussion of the 'argument by analogy' and its flaws.

are beyond other creatures, and presumably in order to do so they require and are equipped with extra 'mental machinery'. However, this does not automatically mean that any shared or more basic processes disappear. Possessing more complex mental processes does not necessarily rule out retaining basic ones. At least for some (simpler) tasks, maybe humans have more than one way of doing things. If this is the case it also follows that there could be more than one way of interpreting what they are doing on the occasions that they are performing such tasks. In some situations an analysis appealing to basic processes might be more than sufficient to account for particular observed behaviours. In other cases, the extra machinery might be necessary for success at some task, and will therefore need to be appealed to in explanation. Behaviours involving the complex processes in their production will require a complex analysis in order to be fully understood and appreciated. In some cases either strategy might work, perhaps with the complex analysis providing a more sophisticated explanation still compatible with the basic analysis.

So, when we say that humans may have more than one way of doing things, we could mean either that when presented with some task they have different ways of doing the same thing – say a complex strategy or a basic strategy. Or we could mean that the things that people do may fall into, say, two categories – those which call upon basic processes and those which rely on complex processes. Or there could be some overlap between these two options, with some situations accounted for exclusively by basic processes, some necessarily by complex processes, and some describable in both ways.

4.5.3 The problem of over-attribution

Let us now return to our initial examples to discuss how these ideas relate to potential over-attribution in interpretation. Take for instance the case of the detective again, and consider his ducking behaviour. The detective could produce this behaviour

even without taking into account the mental state of the cop he wants to avoid.⁸⁴ Alternatively, a calculation of the cop's mental state might have featured explicitly in the motivation for the detective's behaviour. As a normal adult human, the detective is in fact perfectly capable of taking into account the mental states of the policeman. Furthermore, it is likely that if interviewed about his actions he would claim that he hid so that his adversary would not know he was there. While, as a good detective, he is likely to have ducked unreflectively, as a good psychologist, he is likely to provide a detailed explanation of his ducking. As observers, we are equally likely to do so, and, chances are, our explanation will be the same as the one provided by the detective himself.

Now, note that if, as it happened, the basic strategy were the one that had been in operation all along, both we and the detective would in some sense be imputing a layer of complexity to the scenario which was not originally present. However, in this case I would be reluctant to say that such over-attribution makes either the detective or an observer seriously mistaken about the situation. Maybe we can say that even if the target behaviour was produced without reference to higher-order processes, it could have been produced with such reference, and, if it had been, both the actor and the observer would be in agreement about what these were. By contrast, if the higher-level processes were beyond the agent producing the target behaviour (that is, they could not be independently attributed to the agent as part of her cognitive resources), we would have to say that the over-attribution is a mistake.

To relate these observations back to the discussion of *MPP*, whether it is appropriate to attribute a particular theory to a creature is a matter for empirical discovery to be determined by ethological observation of the creature. However, this is not something ordinary individuals have the time, resources or inclination to carry out. Even if they did, it is not clear that they could alter the theory they use except

⁸⁴ For example, resorting to the same type of strategy arguably deployed by chimps to avoid the unwanted attention of other chimps, manifested for instance in their manoeuvres around dominant animals to reach food. See Povinelli & Giambrone (2001); Tomasello (2002) for discussion.

via explicit reasoning. People just apply what they have. So, when humans use their theory to explain some other creature, the problem of over-attribution will arise unless the particular explanation rests on some law basic enough to be present in the less complex theory as well. For example, explanations that rely on (some varieties of) WILLING N (necessities) will probably be a feature of many theories. When explaining creatures with less complex theories, the human theory will work if the generated explanation relies on some law which is shared between the two theories, but will over-generate if it invokes a law whose introduction the creature's behaviour would not justify. By contrast, when the theory is used by humans on other humans it is likely not only to work but also regularly to be correct.

4.5.4 A 'reinterpretation' hypothesis

According to Povinelli & Giambrone (1999), many theorists interested in the study of other minds explicitly or implicitly tend to adopt some form of *argument by analogy* along the following lines: similar overt behaviours are to be explained by similar underlying causes. Apparently, all the way back to Hume, people have reasoned: *when I think x I do y; so when I see an organism doing y, it is thinking x*. However, Povinelli & Giambrone argue, such widespread appeal to the argument by analogy in the study of behaviour is misguided. When we ask how different organisms understand aspects of the physical and social world we should be careful to establish whether their skills are underwritten by the same concepts and principles.

In particular, meticulous experiments have concentrated on chimpanzees' understanding of the behaviour of physical objects – on whether chimpanzees have a folk physics. That chimpanzees sometimes use physical objects as tools is widely interpreted as indicating that they have an understanding of the underlying physical explanations mediating their success, and of the causal structure of events. However, it is also possible, and in fact appears more probable after much testing, that while chimpanzees are highly attuned to the external manifestations of physical events and to statistical causal regularities, they singularly fail to characterize situations in terms

of cause and effect, or indeed to appeal to any unobservable theoretical entity like *force* or *mass* which humans readily deploy in solving physical problems. For example, when presented with two blocks, one functional with squared-off edges and one a sham with rounded edges, chimpanzees are able, by looking at the blocks, to make a crude diagnosis of causal irregularity about why the sham block keeps falling over, and soon stop trying to make it stand up. When the blocks are L-shapes identical in appearance but with hidden weights at different points inside them, so that in the sham case it is impossible to stand the block up without it overbalancing, the chimpanzees never even look at the blocks to diagnose the causal problem and continue to try in vain to stand up both. Small children, by contrast, are rather good at this kind of task (Povinelli 2002).

A similar situation holds in the realm of social understanding. While behaviours that chimps exhibit (and many types of birds and other animals too, for that matter) might be described as *deception* or *manipulation* of others in their group, careful analyses of the tasks they are performing, and testing different explanations of the skills exhibited in performing such tasks, support the hypothesis that whatever principles chimpanzees adopt to interact successfully with each other, they do not call upon mindreading abilities involving representations of mental states that as normal adult humans it is so tempting to ascribe to them. For example, watching the animals' natural 'begging' gesture gives rise to a strong feeling that the chimp is requesting food from the person in front of them. However, when presented with two people with food to hand over, one of whom can see the chimp and one of whom cannot (because her head is turned the other way or she is blindfolded) the chimp picks who to gesture in front of at random.⁸⁵ This suggests that chimpanzees

⁸⁵ Their performance does improve after exposure to many instances of the task. But subtle variations in the experimental set-up led investigators to conclude that chimps learn and apply increasingly complex rules about certain features of the situation, not that to succeed they have grasped that the task is about *seeing* (Povinelli & Eddy 1996).

do not appreciate seeing as a psychological state that is connected to the focus of attention of the individual (Povinelli, Bierschwale & Cech 1999; Povinelli 2000).

This is not to say that chimps are not highly intelligent creatures equipped with mental states and numerous concepts – just that the concepts they are equipped with do not include ones like mental states themselves and other unobservables. On current evidence, chimps most certainly have mental representations; they just don't have concepts of mental representations. The concepts they have and use are based on perception of statistical regularities. They are extremely good at processing statistical regularities and correlations between events, which goes some way towards explaining how it is that they exhibit the (complex, interactive) behaviours they do.⁸⁶

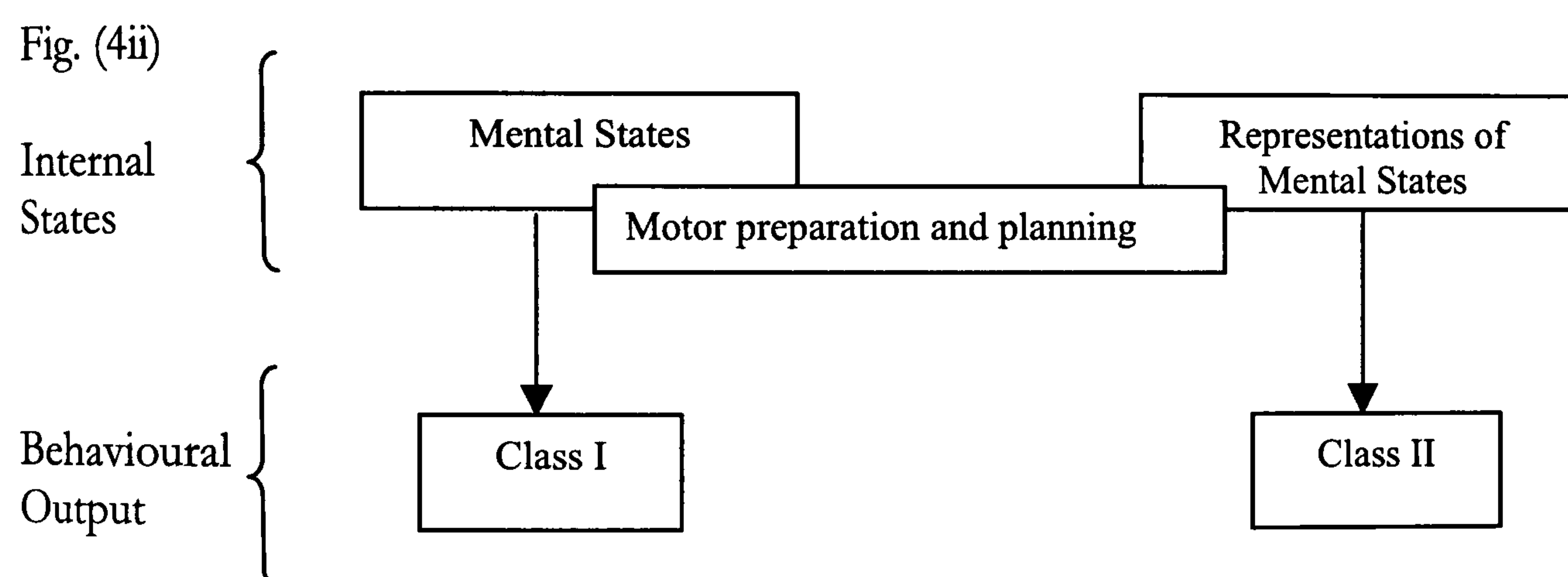
It would be a mistake to think that, because we are talking about understanding behaviour in terms of manifested regularities, a behaviourist account is being proposed. What mediates chimp behaviours is not some kind of stimulus-response process – their behaviour is mediated by complex representations of the world around them. But those representations do not appeal to unobservables.⁸⁷ We can contrast their situation with that of adult humans, who, as we all know, see psychological explanations all over the place (even in situations where they are prepared to acknowledge that it is unlikely that any is in operation – for example, providing intentional ascriptions to moving two-dimensional triangles on a television screen; see Heider & Simmel 1944).

It might be thought that social behaviours manifested by chimps emerge as the result of mindreading. Also known as the 'social intelligence hypothesis', the idea that "theory of mind skills evolved hand-in-hand with increasing social complexity" (Povinelli 2000, p. 60) has been favoured by, among others, Humphrey (1986), Byrne & Whiten (1991) and Baron-Cohen (1995). Instead, Povinelli & Giambrone

⁸⁶ Sensitivity to statistical regularities could still embody principles – just principles about perceptual cues not mental or causal cues (for example, chimpanzees' 'principle' of contact, see Povinelli 2000).

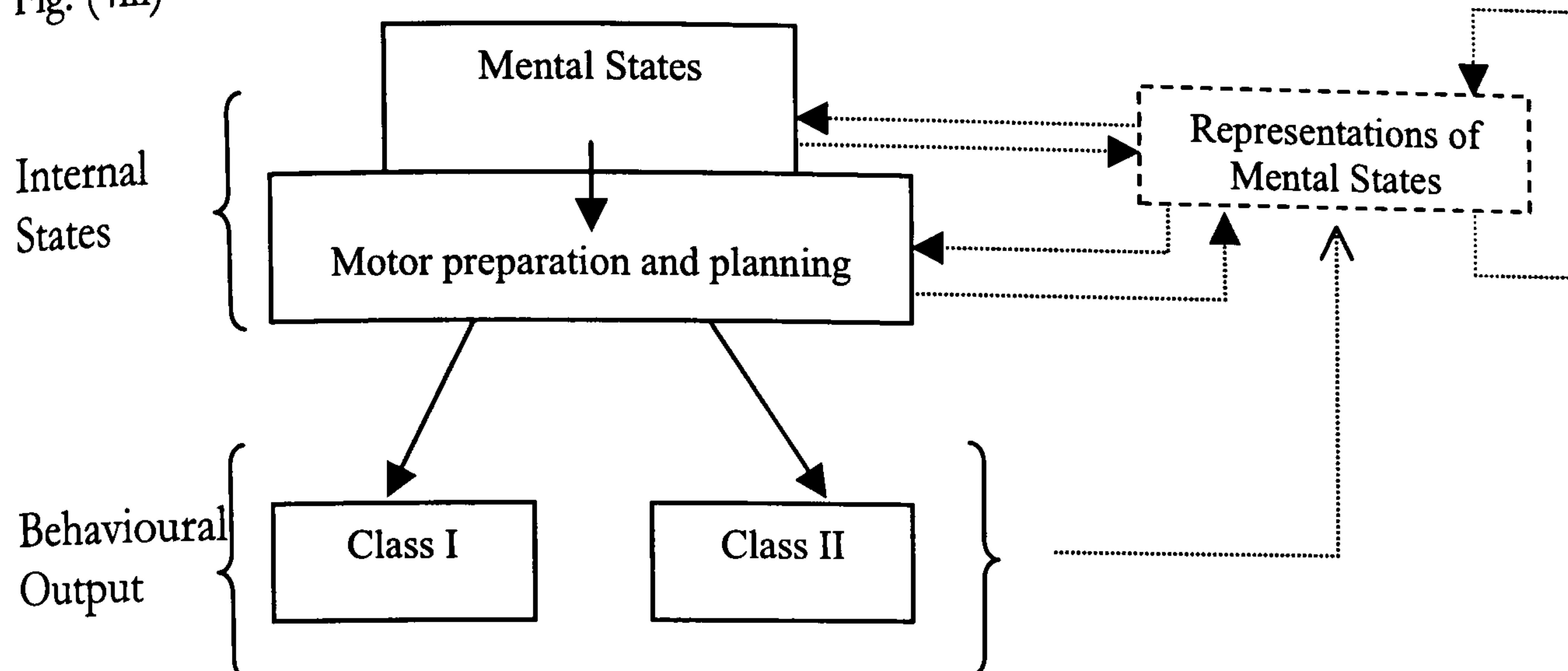
⁸⁷ If we want to talk about behaviourism, it is more like saying that chimps are behaviourist psychologists, but that behaviourism is wrong. Humans, by contrast, are mentalist in their understanding of behaviour.

hypothesize that the opposite is the case: that many of the manifested behaviours humans appeal to second-order mental states to explain existed long before any creature had the means to interpret them mentalistically. These alternative accounts can be illustrated by the following diagrams (due to Povinelli 2002b). In figure (4ii), class I behaviours – things like eating, foraging – are mediated by mental representations, as discussed above. Class II behaviours – ‘social’ behaviours, those usually described as e.g. deception, reconciliation, etc. – are assumed to result from second-order mental states, as per the argument by analogy. By contrast, figure (4iii) illustrates the possibility that first-order mental states could give rise to behaviours that can be so described once second-order mental states are available.



Above, the assumption is that mental states themselves have to be represented in order for the production of class II behaviours to occur. Below, the possibility is put forward that first-order mental states are sufficient for the production of such behaviours. However, humans' psychology (as indicated by the dotted part of the diagram) is such that these behaviours can also be caused by second-order mental states. Furthermore, once the possibility of interpreting behaviours in terms of mental states is allowed, the strategy is applied to all manifestations of behaviours, independently of underlying causes.

Fig. (4iii)



The idea that humans super-impose psychological explanations on ancient behaviour patterns is labelled the *re-interpretation hypothesis* (e.g. Povinelli 2000, chapter 12; Povinelli & Giambrone 2001). Re-interpretation may itself influence and give rise to behaviours, as well as providing, after the event, an analysis of behaviours in the causation of which second-order states were not directly involved. According to the reinterpretation hypothesis, the emergence of mindreading provides new ways of understanding old behaviour patterns, not a sudden increase in behaviour patterns. Of course, once you have started to provide explanations of behaviours, then an explanation itself might become the cause of further behaviour. But note that what behaviours can look like is actually quite limited – what changes is how sophisticated an analysis you are able to provide of them; and how good you are at evaluating whether such an analysis is justified.

So, the point is that it is not necessary that small children or chimps calculate mental states in order to produce behaviours that make it look as if they are calculating mental states (to anyone who can do so).⁸⁸ We might also note that the

⁸⁸ One instance of over-attribution in humans could be that care-givers are inclined to treat small children as if they possessed concepts of mental states regardless of how old they are. Povinelli (2002a) argues that babies have evolved to produce behaviours that exploit adults' re-interpretative tendencies. Atran (2002) argues that the automatic adoption of mindreading schemas is sensible in the face of unknown predators, and further suggests that this principle foreshadows religious notions of omniscience.

theoretical choice in accounts of how creatures come to understand behaviour is not between postulating a system sensitive to behavioural regularities *or* a system sensitive to mental causes – but between postulating a system *just* sensitive to behavioural regularities or a system sensitive to *both* behavioural regularities *and* mental states. Humans as well must have a sensibility towards the statistical regularities of behaviour. Based on these ideas, Povinelli & Giambrone advocate a view of mindreading which is different from that developed here, raising an interesting question about the Baldwin experiments reported above. The issue is whether the action parsing experiments tap into sensitivities dependent on ancient systems or processes which form the basis for further analysis by the psychology faculty.

While it is quite possible that humans have inherited ancient systems, the fact that in humans the EFs and IFs are linked as they are, suggests that whatever residual systems are still around are not involved in this case. However, the possibility that humans have inherited ancient systems may carry implications about whether mechanisms such as Baron-Cohen's Eye-Direction-Detector (Baron-Cohen 1995) are in fact part of the psychology faculty, co-opted by it, or not manifestations of early mindreading at all.

4.6 OUTLINE REVISITED

At this point, it is useful to review the shape of the proposal I am putting forward here, why I think it is on the right track and how it can be contrasted with other positions. We started with the assumption that mindreading is a practice which results in part from some, possibly unconscious, innate body of knowledge or principles. It is not necessary that there be innate knowledge involved in mindreading, but evidence about uniformity of development and universality points to some core innate principles about minds by which cognition in this area is constrained (see also 5.1 below).

I have adapted some proposals from Grice (1975) as a speculative first step in specifying the nature of I-psychology. In particular, I have supposed that humans cognize a psychological theory which features what Grice describes as first-stage internalization of psychological concepts. This theory is part of the cognitive system which realizes the function of planning and executing behaviour and thanks to which an individual is capable of behaviour which bears on the plans of other creatures. I noted that such a system has a ‘comprehension’ as well as a ‘production’ function and that the comprehension function is what provides the parse of social situations in terms of other agents realizing plans. An I-psychology of this type is capable of providing human agents with rich descriptions of social situations. The system is fully recursive, and so situations of indefinite complexity can be readily grasped (modulo the usual limits of capacity). The vernacular-commonsense notions such as *belief* and *desire* are the product of reflection on the output of the situation parser. This reflective common sense may be the product of other faculties of the mind, not necessarily specialized for this domain, but of necessity dependent on and constrained by the actual content of the I-psychology.

One reason for pulling apart what is called upon in mindreading in this way has to do with the fact that people seem to easily and intuitively grasp quite complex social interactions without being able to articulate their grasp of those situations in terms such as ‘thinks’, ‘wants’, which express vernacular-commonsense concepts. I consider more evidence of this in 5.6. Another reason is to allow for variation across cultures as to the kind of vernacular commonsense concepts which arise and variation within cultures as to the levels of sophistication with which individuals develop and deploy these (see 5.1).

In the preceding section (4.5.4) we saw how the innate basis of mindreading might not necessarily include psychological principles. It could be suggested, contrary to my proposal, that the basic components involved in mindreading are inherited ancient mechanisms largely shared by species that do not mindread in the way humans do. These mechanisms would give a first parse of social situations

which could then be re-interpreted thanks to more recently evolved systems which yield an understanding of behaviour in second-order mental state terms. This is essentially Povinelli & Giambrone's position (see Povinelli 2001, p. 241 ff.). In fact this position would be compatible with classical philosophical positions on the status of folk-psychological explanations (see 5.2.1). However, I have suggested that this view cannot account for the way in which 12-18 month old infants form concepts of actions by appealing to underlying intentions, as they appear to do in word-learning; and it leaves unexplained how infants can, but apes cannot, learn by imitation. The developmental and ethological evidence points to humans' parsings of situations being in terms which describe behaviour as action realizing some plan or goal.

In figure (4i) four components were posited as contributing to mindreading abilities. The reasons for drawing attention to each of these in my outline are as follows. There are deep principles that explain our mindreading abilities and we may have only partial (if any) conscious access to them (the I-psychology represented in the action parser). At the same time, we are aware of some concepts that are to do with mental states, and manipulation of these is part of the phenomenology at least – and possibly part of the practice as well on occasion – of understanding behaviour, particularly in reasoned and complex rather than fast and little-considered cases (hence the introspector). Natural language is clearly a factor in how all this is discussed; but natural language resources should not constrain the concepts available in the introspector, as we may well have more mental state concepts than words to express them. Conversely, if there is variety in the concepts employed by the introspector, then the influence of the resources of particular natural languages may be a factor. For both these reasons emphasis was placed on the interface between the introspector and natural language. A further element to take into account is encyclopaedic psychology, where stereotypes and specifically psychological facts are stored. Any of this information may be used to predict and explain behaviour; how much of it is in the Psychology Faculty is to be determined.

One possible query about this picture that might arise on the basis of my initial discussion is whether this account is a kind of *simulation theory*. I don't think so. I discuss why not in more detail in chapter 5. Briefly, my feeling is that simulation is a label which apparently does a lot of work without actually explaining much. To call what I am proposing simulation would be little more than a terminological move. For whatever what I am proposing amounts to, it is not an example of a failure of the so-called 'dominant explanatory strategy in cognitive science', whereby cognitive abilities are accounted for "by appeal to internally represented knowledge structures" (Stich & Nichols 1995, p. 125). What I am proposing does not do away with the need for a 'theory' in this domain, far from it. If anything, what I am proposing underlines the need for the dominant explanatory strategy to be applied properly in this one remaining recalcitrant area – humans' commonsense understanding of psychological matters – which still appears to be in the grip of methodological dualism (see 5.2.2). If what I am proposing counts as simulation, then so does the Chomskyan programme for linguistics, which would be a bizarre result indeed given how the lines in the simulation/theory debate tend to be drawn.

It remains for us to reconsider the vernacular-commonsense concepts (like *belief* and *desire* and *hope* and *expectation*) which so exercise many theorists and are often the starting point for interest in this area. How is it that, in addition to understanding others in terms of the posited internalized theory of acting, we come to think and talk about some of the kinds of states we implicitly appeal to in doing so? And even more, how is it that the vernacular-commonsense concepts end up being the ones they are, and being recognizable across individuals, despite being paramount instances of unobservable entities? In fact, I have suggested that there may be many more vernacular-commonsense-like concepts available to us than we might immediately think, many of them not lexicalised. All these concepts are represented at the level of the introspector. We might then ask why it is that as people reflect on and elaborate their core commonsense understanding the resulting introspective

concepts converge. Perhaps the answer is that people are using the same endowment to refine the concepts of their core common sense. Possible candidates are the science-forming faculty (providing the capacity for inference and metarepresentation), or the language faculty. In other words, as humans, not only do we have in common an I-psychology, we also have in common the reflective machinery on the basis of which we develop concepts which various natural languages encode with words like 'belief'.

Objections to the idea that the SFF might be involved in giving rise to introspective common sense on grounds of universality should not worry us as they legitimately did with regard to whether the science forming faculty could explain how core common sense is acquired (in chapter 3). After all, while basic mindreading judgements are roughly shared by everyone within the same time frame, this is not necessarily the case for more reflective judgements. There probably is a lot of variation in how good people are at appreciating nuances of the mental lives of others and providing explicit characterizations that go beyond immediate recognition of who knows what, where and when. To sum up: humans' understanding of psychology arises on the basis of an LT (H, Ψ) which develops into an I-psychology. This is the core common sense for this domain. Reflective common sense is what is in the introspector for psychology, partially lexicalised in vernacular commonsense terms. Reflective common sense is not underwritten by a specific LT, but may be the result of the application of particular faculties to an LT-determined system.

A reminder may be in order before we go any further. The aim of developing this picture is to separate out various components of mindreading with a view to raising questions about how they fit together. It is not (directly or primarily) an attempt to provide an exhaustive account of mindreading. However, these elements taken together begin to explain how we get cultural and individual variation built on top of reliable underlying systems. In the next chapter I return to the comparison between language and mindreading to argue that differences between them have been overstated.

5

LANGUAGE, MINDREADING AND COMMON SENSE

5.1	EXTENDING THE COMPARISON.....	158
5.1.1	<i>Variation in mindreading abilities.....</i>	<i>159</i>
5.1.2	<i>Further points of comparison with language.....</i>	<i>163</i>
5.2	DISANALOGIES.....	169
5.2.1	<i>Is folk psychology more like folk physics or more like I-language?.....</i>	<i>169</i>
5.2.2	<i>Is knowledge of folk psychology accessible?.....</i>	<i>182</i>
5.3	DEFLATING DUALISM.....	194
5.4	LANGUAGE ACCORDING TO REFLECTIVE COMMON SENSE.....	199
5.5	LANGUAGE, MINDREADING AND COMMON SENSE.....	204

5

LANGUAGE, MINDREADING AND COMMON SENSE

In the first part of this chapter I pursue further the comparison between language and mindreading begun in the last chapter and address possible objections to this undertaking. The second part of the chapter considers commonsense ideas about language and mindreading and their relation to communication.

5.1 EXTENDING THE COMPARISON

Language and mindreading have been compared before; however, I think potential symmetries between them are not generally sufficiently drawn out, that the range of similarities has been underestimated and that alleged differences deserve to be scrutinized. Existing interest in comparisons between language and mindreading has focussed primarily on two questions: the plausibility of appealing to tacit knowledge to explain the abilities at issue in each case, and the status of any theory underlying each ability – whether, that is, the theory is true of the domain it characterises. Generally, language and mindreading are thought to diverge on these points. Before discussing these specific issues below in 5.2 (truth and truth-evaluability in 5.2.1, accessibility in 5.2.2), I would like to extend the comparison beyond its usual confines and touch upon some of the many other areas where considering what we know about language can contribute to our understanding of mindreading.

5.1.1 *Variation in mindreading abilities*

An immediate question is how much and what kind of variation in mindreading do we find? Is variation in commonsense psychological understanding similar to variation between different languages, or is it (even) more superficial than that? As we will see, the variation we find depends on how widely we construe mindreading. In many ways, commonsense psychological understanding is strikingly stable. Behaviour on the basis of mental state attribution exhibits little variation similar to cross-linguistic variation. Standard Sally-Ann tasks are understood in the same way all over the world.⁸⁹ On brief reflection this result conforms to our experience of other people. Going abroad we may desire to learn basic elements of the language of the place we are going; but we neither expect nor encounter wholly novel ways of attributing mental states to others (though it is sensible to be aware of specific cultural traditions and sensitivities in order to appreciate the consequences of certain attributions). Seeing still leads to knowing; false beliefs lead to predictable errors. Silent films with plots dependent on appreciation of the mental lives of their characters require no translation. Mr Bean and Buster Keaton are immediately comprehensible to everyone (even if they are not always amusing to everyone); Tom & Jerry and Roadrunner are shown all over the world without alteration. Spoken films often need dubbing, but their new audiences do not need subtitles in order to understand the *psychological* elements of the plot.

While the basic structure of a situation is likely to be understood in the same way across the board, what might vary cross-culturally is the explicit interpretation assigned to the situation, as this will depend on the detail of what concepts the introspector of an individual in a particular culture and linguistic community has developed. In order to explore the possible variation between commonsense

⁸⁹For discussion and replication of the kind of false-belief tasks used to test mindreading abilities applied to children of different cultures see e.g. Harris (1990); Avis & Harris (1991), Vinden (1996).

psychological concepts at the level of the introspector it could be useful to examine the work of ethnopsychologists.

Ethnopsychologists are social anthropologists who specialise in cataloguing and systematizing the network of psychological words and concepts employed by people from different cultures.⁹⁰ The theoretical background of this kind of ethnopsychology is often far removed from the type of approach to the mind we have been discussing.⁹¹ An example of a typical line of analysis is that if some culture has no word for *anger* then its people also lack the corresponding emotion (Lutz 1988). Stich and Mallon (1999) argue that adherence to this doctrine stems from unacknowledged commitment to certain Lewis-style functionalist ideas about semantics, which they label ‘thick description theory’ and attribute to social constructionists. According to the thick description theory, the meaning and reference of mental state terms, including those for emotions, are determined by the theory in which such terms are embedded. But the terms at issue are everyday mental state terms, part of vernacular common sense, and the theory, “folk psychology”, is allowed to include a very wide range of cultural facts. The result is that being unaware of some connection between the application of a term of vernacular common sense and some culturally determined fact can cast doubt on possession of a concept. So lack of shared cultural tenets will result in lack of shared emotions.⁹²

Whatever is behind ethnopsychology of this sort, according to the picture presented here it is an unfortunate position, as it conflates natural language terms and the individuation of mental states. Instead of ascribing mental states and emotions on the basis of natural language resources, we might prefer to say that emotions are

⁹⁰ For different perspectives see e.g. Rosen (1995); Lillard (1997; 1998); Vinden & Astington (2000).

⁹¹ Indeed this is another example of the prefix ‘ethno’ – like ‘folk’ – applying to more than one topic and associated school of thought, as discussed in chapter 3.

⁹² If this is right about the source of confusion in arguments about the universality of emotions it gives us another reason to distinguish at least between encyclopaedic psychology and core-commonsense psychology. On the universality of emotions see e.g. Ekman (1982); Ekman & Davidson (1994); Griffiths (1997).

universal but how they are lexicalised and what causes them can vary.⁹³ Again, in a full account of mentalizing, these sorts of examples could be discussed in more depth to illustrate the role of the introspector and its interaction with encyclopaedic psychology and memory more generally, although I do not intend to develop this line of thought here.

Nevertheless, however remote from our pursuits many of its conclusions, taken descriptively the field research of ethnopsychologists does provide a lot of raw data about the words people use to talk about various psychological states and the properties they tend to associate with particular mental state concepts.⁹⁴ And it is not only from cross-cultural studies that we can learn about the relation between language and reflective commonsense concepts and underlying knowledge structures. Collecting information about how adults within the same community think and talk about mental state attributions is also an important step in collecting relevant evidence about these relations.⁹⁵

There are other points about variation to address, beyond the issue of cross-cultural differences. For example, we could ask what *variation in endstate* is there among individuals within the same community and *when* do we count that endstate as having been reached? One possible contrast with language that could be argued for is that children of 6 or 7 are already competent speakers and their knowledge of language is comparable with that of adults, while folk psychological ability continues to mature

⁹³ We could say the same thing about ‘obsolete emotions’. It has been claimed (Harré 1986 cited in Stich & Mallon 1999) that *accidie* – dejection or black despair according to the OED, “boredom, dejection or even disgust with fulfilling one’s religious duty” according to Harré – no longer exists. But I don’t think it’s true that “modern people no longer associate any specific emotion with laziness or procrastination in the carrying out of tasks that duty demands” (ibid.). Lots of people feel like that every day at work.

⁹⁴ In the same vein, Stich & Mallon (1999) argue that collaboration between social constructionism and evolutionary psychology, however improbable, might be productive.

⁹⁵ See for example Malle & Knobe (1997, 2001) on the ‘folk concept’ of intentionality.

with age, adults being much better at predicting and explaining behaviour than children. Sterelny (2002) holds a view similar to this, claiming for example that it can take a lifetime for e.g. Papua New Guineans to fully develop their folk biological expertise, and that the same is true for folk psychology, but not for language.

This contrast between linguistic knowledge and commonsense knowledge may be only apparent, and needs to be spelt out in more detail in order to be evaluated. First, let us assume that we are talking about competence rather than performance. Otherwise we could just question the first premise of the argument, i.e., that adults and children have comparable language skills. After all, adults are usually more eloquent and articulate than children due to having had more practice at expressing themselves; but presumably this is not the sense in which adult and child language ability are being held equipollent. We could grant, then, that a reliable endstate (comparable with that of an adult) has been achieved by children in the case of language by the age of 6 or 7, but might also note that there is still room for the addition of, for example, novel lexical items, and that adults have a much larger vocabulary than children. This could be one way in which adults in fact do have greater mastery of language than children independently of performance factors.

Now, a defendant of the Sterelny view could counter that while yes, lexical items continue to be learnt throughout one's lifetime, at no point is the speed of acquisition of such new items comparable to the speed at which children first begin to learn words. Even someone who suddenly has to familiarise themselves with the terminology of a new job is unlikely to have to master more than say 40 new words (and may still have trouble doing so). Children learn at least that many every week or so in the first few years of their lives (see e.g. Bloom 1993, 2000; Barrett 1995).

But who is to say that that is not precisely the situation the natives of Papua New Guinea are in as their folk biological knowledge allegedly increases? In order to sustain a contrast with language, one would have to check *in what way* their folk biological knowledge continues to develop throughout their lifetime – does it qualitatively improve, or just get bigger on exposure to more species but still respect

he same principles as were present from the beginning in categorising these new specimens? If the latter rather than the former is the case for folk biology, then the argument for a contrast with language is weakened.

Even in this case, a further attempt to safeguard a contrast with folk psychology could still be made. Someone could argue that in the case of language (and perhaps as it turns out biology) the endstate expands through addition only of individual items according to structures that are in place from early on, but that in folk psychology what can be added are not just the equivalent of lexical items, as it were; whole new principles can be added, e.g. new views about how political opinions affect behaviour, or about social stereotypes, which increase the power of the whole system and thus the expertise of the adult folk psychologist in predicting and explaining behaviour.

But if this is the way in which the contrast between language and folk psychology with regard to adult and child endstates is maintained, it comes at the cost of having cast the net very widely in individuating what a system that is comparable to I-language might be. And it doesn't argue against the possibility that there is some part of folk psychology that is in place early on and which does not change much after that. In short, what Sterelny's point actually highlights is that different aspects of folk knowledge could be called upon to perform some intuitively individuated task, and we should be clear about which are at issue in making claims about development.

5.1.2 *Further points of comparison with language*

Other questions we might ask include: Is there a *critical period* for acquisition of mindreading abilities? If so, *when* is it? Bear in mind that these questions may have different answers depending on whether we are asking about a critical period for the LT (H, Ψ) or for the introspector. While the development of I-psychology may be a candidate for the kind of process that has a developmental window, the development of the introspector might, in practice, stretch to life-long reflection on mental states and their combinations, yielding, over time, increasingly sophisticated appreciation of

behaviours. Nevertheless, as noted above (in 3.7 and 4.6), we should attempt to distinguish between aspects of the introspector which are present on the basis of limited information across the board, if there are any, and aspects of the introspector which are affected by quantity and quality of experience, and individual skills.

The idea of a critical period is tied up with there being something that has to happen during some period. So, where there is a critical period, is there *specific evidence* to which an individual has to be exposed in order for mindreading abilities to develop intact? If so, *what* is the evidence that a child critically has to be exposed to? For language it is exposure to samples of language. For mindreading, do people just have to see others behaving? Acting on false beliefs? Talking about mental states? Of course, as we have distinguished between full-fledged mindreading abilities and lower-level psychological understanding, we should also ask how these are (differently) affected by what evidence.⁹⁶ There has been some suggestion that children who have siblings (or presumably older playmates) develop mindreading slightly faster than children who don't (Perner, Ruffman & Leekham 1994; see also Lewis et al. 1996). Is it plausible that this environment should facilitate acquisition of mindreading? Of which aspects of mindreading? And why, exactly?

Sterelny (2002) has argued that we have a *wealth of the stimulus* argument against an innateness hypothesis for folk psychology – that children are “soaked” in interpretive environments and there are cultural inventions like narrative stories which often rehearse interpretations to re-inforce folk psychological understanding. But whether the stimulus is ‘wealthy’ *enough* to do away with innateness rather depends on the overall complexity of the system you end up acquiring. Furthermore, poverty of the stimulus arguments in the case of language acquisition are more about poverty of quality than poverty of quantity. The issue is not so much about the *amount* of information available (after all children can be pretty soaked in linguistic

⁹⁶ Siegal (2003) suggests that lack of exposure to talk about mental states results in delayed success on standard false-belief tests (arguing that the performance of congenitally deaf children of non-signing parents is at the same level as that of autistic children, but for different reasons).

environments too), but the fact that the available evidence is *degenerate*, and that *variability* of the stimulus across individuals appears to have little effect on the resulting knowledge structure.

There is also the question of how much explicit teaching can help in the acquisition of various cognitive capacities (what we might call an *irrelevance of the stimulus* argument). Baldwin and her colleagues (2001) raise the issue of whether in interactions with children care-givers underline the completion of actions to try and help children discern one action from another. They cite an unpublished paper (Brand et al. 1999) that considers whether *motionese*, an equivalent of *motherese*, comes into play in helping children to understand the structure of behaviour, and claims that “parents’ actions toward infants tend to include greater repetition than do comparable actions directed towards adults, and infant-directed actions include other modifications (e.g. simplification and increased range of motion) that serve to highlight structure within intentional action” (Baldwin et al. 2001, p. 714). ‘Motherese’ is the name of the kind of baby-talk that parents (think they) address their children with. While it may be a widespread phenomenon among parents, there is little evidence that it makes any difference to the speed of language acquisition of the child.⁹⁷ Similarly, one might conjecture that even if adults do re-inforce their action movements in their behaviour towards children, it does not necessarily follow that children need this input in order to develop their mentalizing skills, or even that children make much use of the exaggerated input when it is available.⁹⁸

One possible contrast between language and mindreading could be in the extent to which development depends on external input and how much it is a matter of pre-determined maturation rather than parameter setting or any other type of slot-filling

⁹⁷ See Gleitman & Newport (1999, p. 3-4). Mothers in fact tend to speak to their children in whole sentences and even where they don’t the child’s learning rate is largely unaffected.

⁹⁸ In fact, research currently under way appears to support my conjecture (Reid 2003).

folkbiological knowledge in adults will inform our understanding of what initial principles might have been in operation and how these can be affected by interaction with different environments (see also Olson & Torrance 1996).

However, some commonsense domains could be closer to the case of language than others, at least in some regards. Certainly some concepts that arise in the study of language seem better suited to some domains than others. Take for instance the distinction between competence and performance. Is performance specific to domains? Could there be language-specific performance and commonsense-specific performance? Could there be performance skills that are specific to particular commonsense domains? There might, for example, be a specific *competence/performance distinction* operating in the case of mindreading. Surian & Leslie (1999) discuss a distinction between competence and performance not just as a methodological point to bear in mind as a factor in evaluating experiments, but as a way of explaining what might be going on in the Sally-Anne task when children appear to have some mentalizing abilities despite failing the task. They argue that children before the age of 3 may have the requisite knowledge (competence) but lack sophistication in applying it (performance).¹⁰⁰

As we have seen, the notions of E-psychology and I-psychology, intended as roughly equivalent to E-language and I-language, may be worth developing. The term ‘I-psychology’ could be useful to distinguish the type of approach that is being developed here from others, and to provide a label for what underlies (some circumscribed set of) mindreading abilities that is free of the history and problems associated with ‘folk psychology’ and ‘theory of mind’. However, in introducing it, we need to decide how much of the proposal outlined above in chapter 4 is included in I-psychology.

¹⁰⁰ Leslie’s idea of a selection processor (SP) could be an example of performance specific to mindreading (Leslie & Polizzi 1998). See also Leslie & German (1995); Bloom & German (2000).

Asking whether there is an equivalent to E-language for psychological understanding could be a rather empty question; one of the points about the notion of E-language is that there is no way of identifying it clearly (other than as not I-language, perhaps).¹⁰¹ As we saw at the beginning of chapter 4, there are a variety of notions related to commonsense psychology that might be counted as E-psychology; but that does not tell us very much about what we are interested in investigating. A more significant question than talking about different ways of construing E-psychology might illustrate is what we should be treating as data for our study of mindreading: (i) sets of behaviours humans are able to parse; (ii) sets of understandings humans produce about behaviours; (iii) sets of judgements about understandings of behaviours.

Thinking about (i), the sets of behaviour we are able to parse, helps to focus on a search for structure to behaviours that could map onto our understandings. Discussing the Baldwin experiments above we noted that, for action parsing to be useful, there would have to be a correlation between what infants take to be perceptual cues for the completion of actions, and the intentional structure of the actions observed – a sort of morphology or phonology for behaviours – just as words and sentences have structural properties that our language systems are sensitive to. (ii) might involve the more inferential aspect of commonsense psychological understanding. (iii) concerns the appropriateness of pairings between behaviours and understandings.

The three possibilities bring out similarities with different problems dealt with by linguistics: (i) could be compared with utterances we are able to parse; (ii) relates to assigning an interpretation to a parsed structure; and (iii) involves assessing the plausibility of an assigned interpretation. There are different kinds of language-related judgements that language users are capable of making. Some are to do with grammaticality, some are to do with meaning, some are to do with acceptability of

¹⁰¹ See Smith (1999, p. 38) on why “the notion of E-language is empty of content and correspondingly unnecessary”.

interpretations.¹⁰² I-language clearly makes a contribution to all of these judgements, but it is debatable whether it exhausts what is called on in delivering particularly the latter kind. We should bear in mind that similar considerations might apply to mindreading judgements.

5.2 DISANALOGIES

We have been comparing language and mindreading, pointing out similarities and where we expect divergences. There are two possible queries about the analogy itself that I have not yet considered. The first is to do with *corrigibility* and the second is to do with *accessibility* of folk psychological knowledge. I will present each one and explain why I do not find the issues they raise problematic.

5.2.1 *Is folk psychology more like I-language or more like folk physics?*

To address this question we need to compare comparisons. Stich (1996, p. 41-42) draws a stark contrast between I-language and folk physics. I propose to consider this comparison in some detail. We can then discuss whether it is more appropriate to compare folk psychology with folk physics or with I-language.

¹⁰² Grammaticality judgements can be independent of meaning and relate solely to the acceptability of a string. Often tested in minimal pairs, an example might be *The cat is black* versus **The is black cat*. Judgements are also made about what certain strings can and cannot mean, what interpretations can be associated with what strings, and about what states of affairs can be described by some string. The data for binding theory, possibilities for co-reference of terms in a string, provide many examples; or contrasts between pairs like *John ate an apple* and *John ate* (which can mean “John ate something”) versus that between *John is too stubborn to talk to Bill* and *John is too stubborn to talk to* (which can’t mean “John is too stubborn to talk to someone” but has to mean “John is too stubborn for any one to talk to him”), see e.g. Chomsky (1986, p. 8). We also make judgements about which interpretation to select out of various possible ones, e.g. in *John can open Bill’s safe. He knows the combination*, where the pronoun is taken to refer to *John*, although there no binding constraints stop ‘he’ referring to *Bill*.

So, how does the theory of folk physics compare with the theory of language? First, Stich notes, there are two ‘theories’ we need to keep track of. A psychologist might study folk physics and characterize as best she can the internally represented principles that govern her subject’s judgements about the behaviour of physical objects. If she does a good job, her theory of the psychology of folk physics will be true. If she doesn’t, her account of folk physics will be false. A separate issue, the more significant one here, is whether folk physics itself, the body of knowledge – or theory – internally represented by the subject and the object of inquiry for the psychologist of folk physics, is true. If folk physics is correctly characterised as roughly corresponding to medieval impetus theory (see chapter 3), then, as it turns out, folk physics is actually false about many physical matters.

Now, take the situation in the case of language, where similarly a body of internally represented knowledge is said to be responsible for individuals’ judgements about e.g. the grammaticality or ungrammaticality of sentences. It is certainly possible that a bad linguist can construct a bad theory, one that fails to characterize correctly the principles that underlie those judgements; we can but hope that better linguists will characterize the subject’s I-language more successfully. But what about the separate issue noted for physics above? Can the subject’s theory, her internally represented grammar, or I-language, be false? Here, says Stich, the case of language and that of folk physics differ. If the internally represented principles of grammar are what determine the subject’s judgements about the grammaticality of a sentence, then the I-language that embodies these principles cannot be wrong about whether or not a sentence is grammatical. The subject’s ‘theory’ of language, according to which sentences are assessed for grammaticality, is true about the language she speaks because the language she speaks is a product of the theory.

Having illustrated these two alternatives, it is interesting to ask whether folk psychology falls in line with folk physics or with language. But before we attempt to answer this question, I would like to make some points about the comparison just

outlined. Returning to folk physics, we might say that if the psychologist's account of the psychology of folk physics is correct, the psychologist will have provided a true characterization of the '*I-physics*' developed and deployed by humans. But note also that the pre-theoretic judgements humans make about physical matters are entailed by their I-physics. So even in the case of physics, I-physics will be true of the judgements individuals make about physical matters in much the same way as I-language is true of the judgements humans make about language. I-physics is constitutive of individuals' commonsense knowledge of physics, even if I-physics is false about physics.

Now, the problem with folk physics is that there are some facts of the matter external to humans – phenomena given by the world and understood by scientific physics – that make individuals' judgements about physics right or wrong. In the case of language, arguably there is no fact of the matter external to me that makes my judgements about (my) language true or false. However, note that there could be a fact of the matter about the truth or falsity of the hypotheses that my I-language unquestionably provides me with when faced with external linguistic strings to parse. Perhaps, although a string sounds like a sentence of English, and my I-language assigns it a structural description accordingly, it was produced by an alien operating according to completely different principles. More realistically, if I start making judgements about someone else's productions, there surely is a fact of the matter about what underlies those productions – whatever underlies the linguistic ability of the speaker – and my theory will only be true about *that* to the extent that we share an I-language; more precisely, to the extent that our I-languages agree on the grammaticality of the string in question.

What makes language different from physics again is that in fact individuals within the same speech community generally *do* share an I-language to a greater or lesser extent, probably making most judgements that individuals might make on the basis of their I-languages correct about the linguistic productions of others. But this is not what is usually at issue for linguists aiming to characterize I-languages. Nor is it

the same question as the one initially considered for language in the comparison above (which was the truth of the theory with regard to *one's own* productions and comprehensions). So the comparison between language and folk physics as drawn by Stich is not being drawn on exactly the same terms. Once it is, we can see that there could be scope for disagreement as to the 'truth' of an I-language as well, depending on what we take it to be supposed to be true *about* – (i) the judgements people make, or (ii) the causes of what people make the judgements about. And the case for truth on the basis of the theory (I-language) entailing the judgements (about grammaticality) only holds for the first issue, i.e. the judgements people make. And for this issue I-physics goes the same way. It is the second issue, i.e. the causes of what people make the judgements about, that leads to the difference between language and physics. But to defend the truth of I-language in *that* case (which I still believe we can) we need more assumptions than are acknowledged, viz. that what the I-language was applied to (some string to be parsed) was caused by something similar enough to it for the assigned structural description to match the original.

Bearing in mind these different questions, we can now try and ask, for each one, whether folk psychology, on a particular interpretation of the term, is true. So, just as for language and folk physics, a theorist of folk psychology might construct a bad theory about what folk psychology amounts to. Let us put the possibility of folk psychology being false in this way to one side for the moment and assume that we have characterised the principles of folk psychology successfully. Call this true theory, ascribable to individuals, I-psychology. Now, can what I-psychology is a theory of be different from how the theory says it is? Well, yes and no. In one sense (issue (i)) it can't. In the same way that I-language is not a theory about (E)language, but a theory of an individual's knowledge of language and therefore incorrigible; so I-psychology is not a theory about (E)psychology, but a theory of an individual's understanding of psychology, and thus determines incorrigibly the judgements she makes about psychological matters and the applicability of psychological terms. In

another sense (issue (ii)), it depends. What is at issue here is whether the hypotheses provided by I-psychology are correct about the underlying causes of the behaviour they describe. On this question, should we say that 'folk psychology' is like I-language or folk physics?¹⁰³

Is it the case that just as there are physical matters that folk physics is ill equipped to deal with, there are psychological matters that folk psychology characterizes mistakenly? It is certainly true that there are psychological matters on which folk psychology is silent – on no one's account of what folk psychology is does it include information about vision employing 2 ½ D-sketches, for example (see Marr 1982). But this information might just be outside the realm of what folk psychology deals with. The question should really be whether folk psychology is right about the bits of psychology that it characterizes. (After all, the problem with folk physics is not so much that there are swathes of facts that it is silent on, e.g. quarks, but that it is wrong even about what it does deal with, e.g. the movement of medium sized objects). So what about those?

At the most general level, this question [A] is about whether there are in fact, as posited by folk psychology, causally efficacious mental states operating in the heads of the agents whose behaviour we explain. A further question [B] is whether the states in operation are the ones folk psychology says they are. And in fact, there are potentially two aspects to this further question. We might ask in general [B₁]: Are the mental states that operate in the heads of behaving agents *of the same kind as*

¹⁰³ Before we answer it, as an aside someone might point out that this latter question is clearly the one that Stich had in mind all along, and that I've taken rather a roundabout route to address it. However, my gripe with how Stich sets out the problem is that you can't conclude that folk physics is false and I-language is true just from the fact that I-language is constitutive of grammaticality judgements whereas folk physics makes the wrong predictions about the trajectory of projectiles. That conflates issue (i) with issue (ii). And issue (ii) is not normally addressed about language, but is what most discussions of folk physics, and by extension folk psychology, focus on. Bringing out this disparity helps to articulate a more detailed and insightful comparison between language and mindreading which otherwise risks getting papered over.

the ones folk psychology appeals to in explaining behaviour? Or we might also ask in particular (if the states appealed to by folk psychology *are* of the same kind as those appealed to by folk psychology) [B₂]: For any specific piece of behaviour on the part of an agent, does folk psychology provide a correct hypothesis about the *combination* of mental states that caused the observed behaviour in that instance? Let us now take up each of these points in turn, as it seems we cannot sensibly ask or decide whether folk psychology is true or false without specifying which one we are addressing.

The first of these questions [A] is about the *underlying causes* of other people's behaviours, rather than about the *understandings* of other people's behaviours that individuals produce on the basis of their folk psychology. As noted above, it is thus a different sort of question from the question with regard to which I-language is standardly assessed for truth. It is however an empirical question. If it turns out, according to science, that behaviour *is* to be explained by mental states, there is a sense in which folk psychology is true; if it doesn't, it's false. Let us assume for the sake of argument that at least this much is true, and proceed.

Whether the states in operation are 'the ones folk psychology says they are' [B₁] is slightly trickier, and the comparison with language will be useful again to explain why. Here we have to ask first what the principles of folk psychology actually *are* before assessing folk psychology for truth about the underlying causes of others' behaviour. To say whether folk psychology is true in this sense we first have to say what folk psychology is. Here the possibility of a bad theorist constructing a bad theory about I-psychology creeps back into the picture. For example, someone might argue that we already *know* what mental states folk psychology says are in operation – they are states of *belief* and *desire* and the like. But, as with language, our pre-theoretic ideas about what underlies our commonsense judgements about psychology may well not tell the whole or the correct story. I-psychology is just as likely to call upon varieties of concepts like PREHENDINGS and JOININGS rather than simply 'beliefs' and 'desires'. Furthermore, the full range of I-psychological competence may call

upon many more concepts than just the ones that have close correlates in pre-theoretic terms. Once we have an idea about the content of folk psychology – a theory of I-psychology – we can evaluate whether the principles of folk psychology correctly characterise the underlying causes of the behaviour they explain. If it turns out that varieties of PREHENDINGS and JOININGS and the like are what is in the heads of others causing their behaviour, then I-psychology will be true.

As suggested by my proposal in chapter 4, I think it is likely that in this sense ‘folk psychology is true’, comparably to language. Now, we still need to address the second aspect of the further question about the truth of folk psychology raised above. The remaining question [B₂] about whether the states in operation in the observed agent are the ones folk psychology says they are, is to do with whether the hypotheses provided by folk psychology are correct about *particular instances* of behaviour. If we are asking whether folk psychology is true in the sense that the explanations it provides about why someone did something are always correct, that intuitively seems unlikely. Even if folk psychology, properly understood as I-psychology, is right about there being PREHENDINGS and JOININGS going on, it doesn’t follow that it is always right about *which* PREHENDINGS and JOININGS are going on. Folk psychology might class some piece of behaviour as an instance of wanting to attract a waiter, for example, when in fact it was just an instance of wanting to repel a fly.¹⁰⁴

Now, to keep exercising all the distinctions drawn so far, note that even in this case, on the original issue (i), what we might call the internal reading of I-psychology being true, we might still want to say that folk psychology is true about the judgement it made. I can’t be wrong about the interpretation I assign to someone’s behaviour. However, to say that folk psychology is always right about the explanations it provides is one thing, but to say that the explanations it provides are always correct is quite another. And to say *that* seems highly implausible. So do we want to go back to

¹⁰⁴ For ease of exposition as well as due to the absence of a worked-out theory of I-psychology I have lapsed into talking about *wanting* rather than kinds of WILLING, as enabled by my introspector.

saying that for the original issue (ii), judgements about the causes of other people's behaviours, folk psychology is false about 'particular instances', i.e. behaviour-explanation pairings? Well that doesn't seem very satisfactory either; we may get some specific instances of behaviour-explanation pairings wrong (due to over-attribution or plain mis-attribution), but the explanations provided by folk psychology do seem to be at least partially right a lot of the time. If that is the case, how should we account for it? What it seems we should be doing for this last question is approaching the problem on a case-by-case basis to try and see exactly where the hypotheses about behaviour provided by folk psychology stand or fall, and whether there is systematicity in the understandings and the misunderstandings it gives rise to. For example, hypotheses about behaviour provided by folk psychology might be wrong due to a mis-match between I-psychologies (compare people speaking different dialects). In such cases problems might be overcome by storing marked exceptions to certain generalizations in certain contexts or for certain individuals. Other ways for folk psychological hypotheses to be wrong about particular instances could be lack of shared context, or lack of sensitivity to relevant contextual features.

At this point a comparison with pragmatics rather than with syntax might be helpful. On a view of pragmatics that sees communication as a form of intentional behaviour, there is a significant overlap between accounts of communication and accounts of mindreading (e.g. Grice 1975b, 1989; Sperber & Wilson 1986/95). It is thus interesting to ask what answers we would give for pragmatics to the types of questions we have just been asking about folk physics, language and folk psychology.

One of the aims for a theory of communication is to explain the fact that humans interpret each other predictably and successfully despite the extent to which the recovered interpretation is underdetermined by the evidence on the basis of which it is recovered (see Carston 2002). However, communication is not always successful, and accounts of communication aim to explain what happens in cases of miscommunication as well. What we want to capture is how interpretations are

recovered and why recovered interpretations are the ones they are; whether the recovered interpretation is the one the speaker intended is a further question. One thing is parsing sentences and assigning them an interpretation; another thing is asking whether that interpretation is what the speaker of the sentence had in mind when the sentence was uttered. While your judgement about how to assign the interpretation follows from your knowledge of language (and pragmatic competence), the matching of the recovered interpretation with the intended one does not. Nevertheless, most of the time the interpretations hearers assign to utterances do tend to match up with the interpretations speakers intended. In other words, you can't be wrong about the interpretation judgement you make about an utterance, but in principle your hypothesis about the speaker's intended meaning could be wrong. What is interesting is that it usually isn't wrong, and when it is wrong, it's wrong for predictable reasons.

The predictable reasons tend to be ones about mismatching initial assumptions. And mismatching initial assumptions go a long way towards explaining folk psychological misunderstandings as well as communicative misunderstandings. What is special about the case of communication is that it is a form of ostensive behaviour which arguably licenses a particular level of expectation about the recoverability of an intended interpretation (Sperber & Wilson 1986/95). No such *specific* expectation is justified when merely watching someone else behave independently of the observer. However, that is not to say either that no levels of expectation at all exist in the case of non-communicative behaviour, or that none of these levels might in some sense be justified. These are open questions. For example, as noted in 4.4, there is a difference, intuitively at least, between being an invisible observer of an individual's activities and being a present but passive observer of those same activities. Further, perhaps specifiable, differences apply to other situations, e.g. between being an invisible, passive, or active observer of non-communicative interactions between two or more people.

Having looked at the different ways in which we might talk about folk psychology being true or false, let us sum up the position I have been defending. I have argued that folk psychology can be true in similar ways to I-language. It is the case both that I-psychology entails the judgements individuals make about psychological matters, and that I-psychological judgements are correct a lot of the time. In a general sense judgements are correct about the states they posit because they are judgements about the minds of conspecifics. Furthermore judgements tend to be correct about particular cases. So linguistic knowledge and commonsense psychological knowledge are true in an 'external' sense as well, unlike commonsense knowledge of e.g. physics and biology. As Fodor puts it for language:

What normally makes one's contingent beliefs reliably true is that they are formed by processes that are sensitive to the way that the world contingently is. But, in special cases like language, what makes one's innate, contingent beliefs true is that they are about the minds of creatures whose innate cognitive capacities are determined by the same genetic endowment that determines one's own.

(Fodor 2000b, p. 95)

and for commonsense psychology:

[In the case of knowledge of language] the facts that make one's innate beliefs true aren't, as it were, ontologically prior to the beliefs themselves. Likewise, perhaps, for the innate intentional theory of mind that many of us nativists think that people are probably genotypically endowed with. The same endowment that determines my innate theory of how your mind works also determines that your mind works the way that my innate theory of your mind says it does. And, of course, vice versa.

(ibid., p. 97)

These points relate back to an earlier one about the significance of parsing behaviours and the correlations between action boundaries and underlying intentions. Perhaps one could argue that just as sentences have certain structural properties because they are a result of a human linguistic system, behaviours have particular structural

properties because they are a product of our psychological system. Although language and folk psychology fall together on this issue, the case for language is stronger than that for folk psychology. A difference for psychology is that the human system for interpreting behaviour can be applied interestingly, even if mistakenly, to situations that are not the product of human psychological systems. We treat many situations as if they resulted from psychological systems similar to our own, just as we would still parse sentences generated by robots programmed with principles that go against UG but happen to turn out as intelligible. But there are more instances of non-I-psychologically governed behaviour than there are of non I-linguistically governed language. In fact, there probably aren't that many cases of non-I-linguistically governed language. And language principles are more specific than psychological principles, so even if over-applied there is little they can do with a stimulus unless it happens to be linguistic, whereas psychology principles over-generalise much more effectively. The result of these kinds of differences between language and folk psychology is that there will be more cases where the hypotheses provided by folk psychology (about other people (other creatures; other things?)) are 'incorrect', than there will be cases where the hypotheses provided by I-language (about other humans' linguistic behaviour) are 'incorrect'.

These ideas about similarities between mental processes helping to account for success in understanding others should not be confused with simulation accounts of folk psychological abilities. Broadly speaking, simulation theory holds that there is no need to posit a body of knowledge about psychological states and their relation to behaviour to account for the fact that humans can predict and explain others. Instead, it is suggested, it is sufficient to run one's own reasoning processes 'off-line' and attribute the outcome of such a process to the person one is seeking to understand (Davies 1994; Heal 1994; Ravenscroft 2003). I do not intend to go into the details of the theory versus simulation debate about folk psychology here (see papers in Davies & Stone 1995a, 1995b and Carruthers & Smith 1996). However, a

thought experiment suggested by Harris (1995, p. 210-211) is useful to illustrate why I don't think what I am proposing amounts to simulation. Imagine you are told that as part of a psycholinguistic experiment subjects were given a list of sentences of English and asked to pick out which sentences they found grammatical. You are then asked to look at the sentences yourself and predict which ones were chosen. The intuition is supposed to be that you will be quite accurate in your prediction of the results of the experiment. Harris uses this example to argue that it is unlikely that you have a theory about the linguistic knowledge of other people and so to make your predictions you just run your own mental processes off-line.

But crucially, note that without a *theory* of language you would not be able to make any prediction at all – so in fact this account does not do away with the need for a theory. To the extent that you perform as imagined on this task, you are actually *processing* the sentences that were given to the original subjects, not *simulating* doing so. Perhaps this just means that Harris would have to concede that the person in his thought experiment would have to rely on “theory-driven” simulation rather than just “process-driven” simulation to carry out the requested task (see Goldman 1995, p. 85). However, to the extent that performance on the requested task is thought to be guided by just theory-driven simulation, we might actually want to question the intuited results of the thought experiment – for it is quite likely that, depending on the sentences chosen of course, variation in idiolects/dialects could cause the judgements the individual makes in the second part of the thought experiment to be rather different from the judgements of the original informants. Furthermore, even in the slightly unusual situation of providing judgements about other people's grammaticality judgements, it is possible that you supplement your knowledge base to take into account information you think will be relevant for your prediction (e.g. age of subjects, whether they are native speakers, what part of the country they are from; see also Smith & Wilson 1979, p. 200). Certainly, when it comes to predicting other people's behaviour more generally, we are not strictly limited to our own experience. So, while it may be that ‘simulation’ in some sense

has a role to play in accounting for some of the range of what are pre-theoretically labelled ‘folk psychological’ abilities, on its own it does little work without appeal to some body of knowledge, and even then it owes an account of how adjustments are made across individuals.

Another doctrine with which my position should not be confused is eliminativism. Eliminativists believe that there are no such things as mental states (see e.g. Churchland 1981). They believe that folk psychology, like folk physics, is a radically false theory, and that therefore the theoretical entities it posits, beliefs and desires and the like, do not exist (and, so, that the behaviour of other people that we explain by appealing to these entities in fact is presumably caused by something else entirely).

By contrast, I am saying that I-psychology is a true theory, but that the states it posits are not necessarily beliefs and desires, despite our pre-theoretic intuitions about what underlies our ability to explain the behaviour of others. While I am questioning the status of *beliefs* and *desires*, what I am questioning is whether they are the right states to invoke when we aim to construct an account of the content of I-psychology. I am not questioning whether states like them (representational states with causal powers – states with some of the properties ascribed to beliefs and desires by common sense) actually exist. According to my position, I-psychology may not be exactly what we think it is, but whatever it *is*, it’s true. More precisely, it is constitutively true of one’s own judgements; it is contingently true of others’ behaviour, depending on similarities in I-psychologies across individuals.¹⁰⁵ For me, folk psychology as construed by eliminativists (the platitudinous framework surrounding *belief* and *desire*) is not the core of what humans use to understand

¹⁰⁵ Someone might contend that even when I-psychologies are the same, a given parse of a situation may not be contingently true (e.g. a person thinking about someone explaining someone else’s behaviour). As discussed by Jacob (1997), Dretske holds that the information content of propositional attitudes but not that of innate representations is causally efficacious. For a defence of the causal power of the information content of innate representations see Jacob (1997, p. 265-267).

behaviour, it is how we might reflect on and talk about what we use to understand behaviour (see 5.2.2 below).

Having considered some of the issues surrounding debates about folk psychology and its status (for more, see e.g. Greenwood 1991), let us return to the position of the theorist attempting to construct an account of the principles underlying humans' commonsense understanding of psychological matters – we want to gain some idea of what folk psychology amounts to that goes beyond pre-theoretic statements about belief and desire, that is, we want to characterize core commonsense psychology. At this point we might ask, what then is the status of the *folk* (in the sense of vernacular commonsense) concept of belief with regard to *this* theory? I think the vernacular commonsense concept of belief is a reflection of the terms in which we see the world thanks to our core common sense, and may point towards a category that features in our commonsense understanding of behaviour. However, as with other concepts of this kind, not all the properties associated with belief in everyday thought and discourse will necessarily be shared with core commonsense categories that play a role in I-psychology; almost certainly many of the properties associated with the concept of belief in technical philosophy are unlikely to feature in many of the categories (core) common sense employs. The vernacular commonsense concept of word perhaps points towards a category that features in linguistic capacities, in a similar way. To spell out this suggestion, we need to consider the other issue where language and mindreading have been thought to diverge.

5.2.2 *Is knowledge of folk psychology accessible?*

According to the proposal being put forward in this thesis, mindreading abilities depend in part on possession of a body of knowledge (in this case about psychological states and their relation to behaviour). This approach is the one familiar from linguistics, and has even been labelled “the dominant explanatory strategy” in cognitive science (Stich & Nichols 1995, p. 123). But it has been claimed

that it is inappropriate to carry over this approach to the domain of folk psychology. A clear statement of what the worry is supposed to be comes from Davies & Stone's introduction to the debates surrounding folk psychology. Because the anthologies Davies & Stone compiled in the mid-90s were considered the state of the art on these topics for a long time, and are still referred back to frequently today, and because of the status of the authors in this area, I take it that the concern is authoritatively formulated. As I find this concern extremely puzzling for reasons I will explicate shortly, I have preferred to quote extensively rather than paraphrase, to avoid the possibility of misrepresenting what I am arguing against. Davies & Stone say:

An indication of the theoretical depth of the explanatory strategy followed in linguistics is provided by the rich and sophisticated structure that grammars have had to be given in order to account for the linguistic data. The complexity and idiosyncrasy of the speaker/hearer's task seems to rule out putative alternative explanations, such as explanations in terms of general learning. *But, in what sense would a folk-psychological theory show comparable richness and sophistication? Here there seems to be a disanalogy with linguistics. For folk psychology is often thought to have a simple and homely content (hence the epithet 'folk'). Once its principles are articulated, they seem to be obvious. (Contrast the principles of generative grammar).*

(Davies & Stone 1995a, p. 9, my emphasis)

As we shall see, on reflection this quote is more indicative of conflicting assumptions than of a failure of the dominant explanatory strategy. There seem to me to be three possible strands to the alleged disanalogy between knowledge of folk psychology and knowledge of language. Someone might try to uphold a contrast between theories of language and theories of folk psychology with regard to: (1) the principles of the theory; (2) the kinds of concepts that feature in the principles; and (3) the overall sophistication of the theory. These strands are clearly inter-related; I will take them in turn, providing more textual back-up to the points I want to make where appropriate. Briefly, I doubt that serious contrasts for any of these strands can be maintained and will show how the contrast in each case can be turned around; rather than a disanalogy between linguistic knowledge and folk psychological knowledge, I

think there is a disparity between which aspects of the two have actually been compared, and that this mismatch has led to confusion.

To start with, it is not clear *a priori* why folk psychological theory should lack the “richness and sophistication” of linguistic theory. Why not expect that, if anything, the former could well be more complicated than the latter, as folk psychology is perhaps less circumscribed and more pervasive than language? From the quote, it seems as if the motivation for saying that folk psychology is simpler than language is to be found in the nature of the principles that are *pre-supposed* to underlie folk psychological abilities. People think that folk psychology is straightforward because the kind of principles that are usually cited feel intuitively plausible. Here are some examples from a variety of sources:

When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so mental responses.

(Lewis 1972, p. 256)

Persons who are angry tend to frown.

Persons who are angry tend to be impatient.

Persons denied fluids for some time tend to feel thirst.

Persons who feel thirst tend to desire drinkable fluids.

Persons who want that *P*, and believe that *Q* would be sufficient to bring about *P*, and have no conflicting wants or preferred strategies will try to bring it about that *Q*.

(Churchland 1988, p. 58-9)

When a normal person is looking at a traffic light which changes from red to green, she usually comes to believe that it has changed from red to green.

If a person believes that all scorpions are poisonous, and if she comes to believe that Henry’s pet is a scorpion, then she will typically come to believe that Henry’s pet is poisonous.

(Stich & Ravenscroft 1996 p. 126)

If *x* is *y*’s rival, then *x* prefers *y*’s discomfiture, all else being equal.

People generally do what they say they will do.

(Fodor 1987, p. 7, 8)

Now, however, note that someone (perhaps an imaginary linguist called *Dlewis*) could readily construct similar platitudes about language. Here's an immediate sample:

Words have meanings.

Things have names.

People generally mean the same things by the same names.

The order of words matters.

Words make systematic contributions to the meaning of sentences.

There are different kinds of words with different functions.

Action words can be tensed.

Words are organised into phrases.

Sentences are uttered in contexts.

The point about these platitudes about language is not so much that they are wrong – most of them are fine as far as they go – the problem with them is that they are hardly sufficient to capture the sophistication of the linguistic abilities of speaker/hearers. I would claim that just as these platitudes about language are barely even descriptively adequate, so the alleged folk psychological platitudes that have been assumed without argument to constitute the content of an individual's knowledge of psychological matters are also barely descriptively adequate.

It is certainly true that when it comes to suggesting plausible candidate laws for folk psychology everyone starts out all right, but tends to stop as soon as it gets hard, appealing to the fact that the general idea is clear from a few examples. That's one of the reasons that Grice's proposal was so interesting – it was an unusually methodical attempt to derive and justify the kinds of laws that folk psychological theory might rely on. And it is noticeable that in the process of refining his first attempts, even what were initially simple laws become more sophisticated very quickly. And as such laws become more fleshed out, they become less immediately recognisable, or platitudinous. As a corollary, the theoretical concepts they contain become less familiar too. This observation brings us to the second strand of the alleged contrast. Consider Davies & Stone again:

One aspect of the platitudinous nature of these principles is that *the concepts that are used in them are not at all remote from the concepts that we ordinarily deploy in our folk-psychological judgements*. As we learn to engage in folk-psychological practice, so also we learn the concepts of belief, desire, preference, trying, pain, and so on. But, as we learn to engage in linguistic practice – to produce and understand sentences – we do not normally learn the concepts – of phrase structure, head, complement, theta-role, c-command, and so on – that figure in the principles of generative grammar. Likewise, the platitudinous principles of folk psychology are not difficult to understand or to accept, for someone immersed in the practice. But the principles of a generative grammar are understood only by theoretical linguists, and may be highly unobvious, even to experts.

(ibid., p. 10, my emphasis)

To parody the claims this quote makes about folk psychology, someone could just as easily say that “as we learn to engage in linguistic practice we come to learn the concepts of *word*, and *meaning*, and *name*, and *sentence*, and *utterance* (and all the other concepts that appear in platitudes about language that we are readily capable of formulating)”. But no-one would thus conclude that such platitudes are what specify our knowledge of language or that such concepts are those we deploy. Indeed, we might also consider how “deploy” is used in the last quote, emphasised above. How do we know, or why should we believe, that the concepts featured in the platitudes are the concepts we deploy in our folk psychological judgements? They may be concepts that can feature in verbal (or internal) reports of our judgements (although they by no means have to – it is doubtful whether non-philosophically minded individuals reflect on behaviour explanations in terms of people *believing* and *desiring* instead of just *thinking* and *wanting*). But then the concepts of *word* and *sentence* are equally likely to be what feature in verbal reports about linguistic judgements, if linguists ever bothered to collect that sort of data from their informants. In short, a clear line of response to this sort of argument seems to be emerging. Just because some law is a platitude, it is not therefore automatically a candidate for an

explanatorily adequate theory of knowledge of folk psychology, any more than platitudes about language are plausibly explanatorily adequate with regard to knowledge of language.

In the light of these observations it seems unjustifiable to claim a difference between the theory of folk psychology and the theory of language with regard to the complexity of the overall theory. The theory for folk psychology could be at least as involved as that for language. But there is a final quote to be considered about this, strand (3) of the possible contrast:

In the case of linguistics, the term ‘theory’ is indicative of the articulation of the body of knowledge into modules of grammar (X-bar theory, theta theory, case theory, binding theory, and so on), and of the complexity of the derivations of structural descriptions for particular sentences from highly general principles. *In the case of folk psychology, there is not such evident promise of articulation and complexity.*

(ibid., p. 12, my emphasis)

Again, the problem seems to be that, in effect, the principles often cited as examples of folk psychological platitudes amount to *folky* folk psychology, that is, reflective commonsense ideas about the content of commonsense psychology. As soon as this mismatch is recognised then the route to constructing an explanatorily adequate account of knowledge of folk psychology, which may well need to posit concepts and principles which feel unfamiliar to ‘ordinary folk’, and go well beyond the scope of recognisable platitudes, is cleared.

If this view is right, there is a striking parallel between constraints imposed by philosophers today on accounts of folk psychology, and constraints imposed in the past by philosophers on accounts of language. Indeed, it is ironic that research into folk psychology should be held hostage in this way at the same time as linguistic research is favourably held up as an example. In his introduction to a collection of seminal papers in the philosophy of language, Ludlow (1997) notes happily that “philosophy of language, or at least a core part of it, has matured to the point where

it is now being spun off into linguistic theory [...] following the tradition of other branches of philosophy that have been extruded into the natural sciences: physics, biology, and perhaps most recently, cognitive psychology.” (p. xiii). Davies & Stone surely count among the many philosophers who now profess to endorse this trend. Ludlow goes on to acknowledge that:

there is admittedly a great deal of resistance to this view. Some philosophers seem reluctant to loosen their grip on the philosophy of language, or even admit that empirical evidence from linguistics might be relevant. *The contrast with the situation in the philosophy of mind is marked.* Few philosophers of mind balk at the idea that what they are studying might have something to do with empirical research on vision or mental imagery or even the neuroanatomy of the brain. Yet philosophers of language continue to resist arguments that appeal to linguistic evidence.

(ibid., p. xiii, my emphasis)

Ludlow finds this tradition mysterious, but puts forward a tentative explanation:

I’m not sure why the situation is like this. Perhaps philosophers are willing to defer to the sciences when it involves a domain that they cannot see (like the micro level in physics or the activity of the mind/brain in cognitive science) but are unwilling to be deferential when it involves a domain with which they have been fluent since childhood – language. If this is the subtext, then the fallacy is pretty obvious. Philosophers have likewise been ‘fluent’ in thinking since childhood, *but the best philosophers of mind do not thereby conclude that they are experts on the structure of their cognitive architecture.*

(ibid., p. xiii-xiv, my emphasis)

What I have pointed out is that, despite Ludlow’s optimism, when it comes to a particular area of philosophy of mind, the study of commonsense ideas about psychology, even philosophers of mind who are otherwise well versed in the ways of the scientific study of e.g. language or perception are reluctant to let go of their pre-theoretic ideas about the content of folk psychology. Perhaps exceptional fluency in

understanding others leads them to believe they are indeed experts on the structure of that part of their cognitive architecture.

So, I have argued that there is a disanalogy of sorts, but that it is a disanalogy between what is being compared, not between the purported objects of comparison. When these are properly aligned, knowledge of language and knowledge of folk psychology turn out to be much more similar than has been thought. The unacknowledged discrepancy helps to illuminate why knowledge of language is widely agreed to be tacit, unlike knowledge of folk psychology.¹⁰⁶ If, as I have been arguing, philosophers are in fact concerned with discussing ‘folky’ knowledge of folk psychology (i.e. reflective commonsense elaborations of core commonsense understanding), then, just like folky knowledge of language, it is hardly surprising that some of what they are talking about is accessible in a way that most linguistic knowledge is not. We need to pair up the right things to ask whether there is a difference with regard to either of them being tacit. I will compare reflective commonsense knowledge about mindreading with reflective commonsense knowledge about language in the next sections; first let us consider a further issue about the tacitness or otherwise of what we are interested in.

In general, even if there were a difference with regard to tacitness between comparable elements of linguistic and folk-psychological knowledge, not much would follow from that according to Chomsky, for whom the difference between tacit and ordinary knowledge lies merely in accessibility to consciousness. For instance, John knows that he cannot say “I kissed passionately Mary.” As linguists, on the assumption that we understand the theoretical terms we are using and so are clear what it is we are claiming, we can say John knows that R, where R is some linguistic rule from which this piece of propositional knowledge of John’s about language follows. R might be something like: “verbs cannot be separated from their objects by

¹⁰⁶ Maibom (1999) for example has argued that philosophical construals of folk psychological knowledge do not fit any account of tacit knowledge that could still count as tacit.

adverbs.” In fact, rule R might be a consequence of the rule R¹: “Case assignment in English observes a strict adjacency requirement.” (See Chomsky 1986, p. 266). More precisely then, we can say that John knows that R¹. As John is no linguist, he does not know that R¹ *holds*. He just knows that R¹.

If this ascription of knowledge to John does not concord with pre-theoretic, commonsense understanding of the concept *knowledge*, we might prefer to say that John cognizes that R¹, where *cognize* “may abstract from certain features of the term ‘know’ in informal usage [...] Cognizing would appear to have all the properties of knowledge in the ordinary sense of the term, apart, perhaps, from accessibility to consciousness. We might say that ‘cognization’ is unconscious or tacit or implicit knowledge.” (ibid., p. 268-269).¹⁰⁷ Chomsky further holds that this difference is not one of theoretical significance, expecting that “conscious beliefs will form a scattered and probably uninteresting subpart of the full cognitive structure.” (1975, p. 163).

However, an intuitive contrast between states of tacit knowledge and other propositional attitude states has been argued for by philosophers. Davies (1989) points out that we are normally reluctant to ascribe beliefs to individuals unless they “grasp” the constituent concepts of the belief. As “most ordinary speakers have no grasp at all upon the concepts of linguistic theory” (ibid., p. 135), Davies concludes that there is a requirement of “conceptualisation” on the part of the individual (as opposed to the theorist) for “ordinary” propositional attitude states that is absent in the case of states of tacit knowledge, and thus that there is a difference between these two types of states that goes beyond just accessibility to consciousness.¹⁰⁸

¹⁰⁷ We should perhaps note that a broad notion of rule seems to be intended in these passages, one covering principles and parameters and redescription of the consequences of a rule. This is important because Chomsky would otherwise be committed to propositional knowledge of computational rules, which he does seem to have denied elsewhere (see Chomsky 1980, p. 54-55).

¹⁰⁸ The conceptualisation requirement is met by concepts in the ascribed belief satisfying the Generality Constraint (see Evans 1982, p. 104), whereby for a proposition F(a), the individual is able to conceive of *a* having property G, or H, or I, etc.; and of *b*, *c*, or *d* as being F.

For the sake of discussion we can grant this difference for a moment, but it is still the case that, as argued above, the potential for tacitness of knowledge of folk psychology mirrors the potential for tacitness of knowledge of language, once we have aligned what it is we are comparing. It is easy to imagine a parallel situation to that of language for folk psychology; perhaps most ordinary mindreaders have no “grasp” (in the sense specified by Davies) of many of the concepts that occur in their I-psychology, particularly if folk psychology actually contains laws like those suggested by Grice. Ordinary mindreaders may well fail to conceptualise all their JOININGS of necessities or PREHENDINGS of action types; nevertheless tacit knowledge of laws containing these concepts may need to be appealed to in accounting for mindreading.

The specification required even for a relatively simple law relating an organism to food consumption takes up three whole pages in Grice’s framework (see Grice 1975a, p. 135-137). By contrast Lewis’s scant three line attempt (above) is much less sophisticated, and in fact is reliant on the implicit knowledge and understanding of the reader to supplement the detail of the postulate in order that it be intelligible. Of course, length and complexity of derivation are not necessarily an indication of a successful theory, but it does seem plausible that the often cited platitudes are instantiations of much more abstract laws which as individuals we have difficulty bringing to mind and which contain theoretical concepts we do not grasp in the way required by Davies, and so must know tacitly. It is interesting to note that Lewis himself changed his mind about the status of folk psychological generalizations:

We have a very extensive shared understanding of how we work mentally. Think of it as a theory: folk psychology. It is common knowledge among us; *but it is tacit, as our grammatical knowledge is*. We can tell which particular predictions and explanations conform to its principles, but we cannot expound those principles systematically. (Pace Lewis, 1972, p. 256, *eliciting the general principles of folk psychology is no mere matter of gathering platitudes*).

(Lewis 1994, p. 416, my emphasis)

from consciousness.” (Chomsky 1975, p. 164-165, my emphasis).¹¹⁰ But there is no suggestion that anything special about the rules themselves would follow from their being introspectively accessible. According to the second part of the earlier quote, all that would change if suddenly you were to become aware of the principles that determine your linguistic ability, is just that ordinary people would be more prepared to ascribe *knowledge* to you in the everyday sense of the term; the becoming aware does not *itself* otherwise change either the content of the principles or the behaviour you produce on the basis of them. If we imagine that the facts were different and that individuals *were* able, by introspection, to see the very rules they employ (rather than just bits and pieces or approximations of them), we do not expect the rules’ being, or becoming, consciously accessible to have a discernible effect on the linguistic behaviour of the people for whom they are or become conscious – the very same rules would still be what govern the assignment of structural descriptions to sentences.

For folk psychology (mindreading) it is also an empirical question whether the rules we employ are available introspectively, but *that* is not a question that can be resolved by introspection – you have to do the science in this case too to establish whether there is a correlation between what is in operation and what is available introspectively. So it is an open question whether the states appealed to by humans’ commonsense psychology are the ones that humans are reflectively aware of. However, even if it turns out that the principles of folk psychology (I-psychology) are readily recognised or called to mind in a way that the principles of I-language are not, arguably that would still be of little significance. Suppose it turns out that Grice’s laws are in fact overly complex and it is established empirically that the actual laws in operation in folk psychology are mostly the ones we think they are pre-theoretically.

¹¹⁰ Examples of cognized facts an English speaker is aware of are “that necessarily bachelors are unmarried and that ‘is’ goes with singular subjects.” (Chomsky 1975, p. 165). Cognized facts the speaker is unaware of (in the theoretical framework of the time) are for example the relative ordering of specific rules for English and the rules of universal grammar underlying language acquisition, like facts about transformations being structure-dependent.

Or suppose it turns out that quite complicated Grice-like laws are in operation but somehow we are, or we can become, conscious of them. Why should conscious access to the laws of folk psychology matter any more than conscious access to the rules of language would? The same laws would be in operation regardless.¹¹¹

My answer would be that it shouldn't matter whether such laws are known tacitly or explicitly *unless the very fact of becoming conscious or of being conscious of folk psychological knowledge plays a causal role in behaviour*. And in fact the same goes for language. So, one thing that might change in a world where we walked around transparently aware of our linguistic endowment is that people might reflect on aspects of their linguistic system which they don't currently wonder about; thus conscious access to linguistic information could affect behaviour. Indeed, we don't need to imagine an entirely counterfactual situation to see how this would work because some kind of linguistic knowledge *is* reflectively available, and notions like *word*, once they have been conceptualised, probably do play a role in behaviour just like any other concept can; in a similar vein, so might concepts of *intention*. This looks like a valid and interesting reason why conscious access to some aspects of folk psychological knowledge could be important.

5.3 DEFLATING DUALISM

There is another reason why the accessibility of folk psychological knowledge may have been thought to matter so much. It is worth considering what it is briefly just before we move on, as it highlights the need to be clear about what is supposed to be at issue for different accounts of folk psychology. The following quote from Gordon (1995) illustrates the problem:

¹¹¹ Also, the possibility that we might be able to call some of this knowledge to mind, and even the possibility that what we might call to mind in the case of folk psychology could be closer to what is in operation than what we might call to mind in the case of language, is irrelevant as long as you don't *have* to call it to mind in order to perform whatever task is at hand.

Theory theorists generally assume that the laws of folk psychology can be formulated in terms of the common-sense mental vocabulary. They *must* assume this if they hold, as most seem to, that the common-sense theory implicitly *defines* these terms. Generative grammarians, on the other hand, have no such compunctions about introducing a technical vocabulary. They want to explain common-sense intuitions of grammaticality, but they are not constrained to do so by way of a *common-sense grammatical vocabulary*. The reason is simple. They don't claim that what explains our grammatical competence is *folk grammar* – the principles that underlie common-sense notions of grammar and common-sense explanations of why one word string is grammatical and another not. But theory theorists *do* generally make the corresponding claim: that what explains our competence in predicting behaviour is folk psychology – the principles that underlie our mental attributions and explanations of behaviour.

(Gordon 1995, p. 177, emphasis in original)

Here we have a further example of a disanalogy drawn between language and folk psychology that is fuelled by the potential for mismatching what is being compared, as noted above. We also have a possible explanation for the confusion. A tension results from the fact that while folk psychology is understood to be what *underlies* mindreading abilities (in this sense folk psychology is like I-psychology and is comparable to I-language), at the same time folk psychology is *required* to contain everyday notions (which makes it more like *folky* folk psychology and indeed comparable to a folky grammar). Apparently, this methodologically unnaturalistic situation arises because of independently motivated philosophical concerns. If you expect folk psychology to provide the meaning of words like 'belief' and 'desire' then you will construct your theory of folk psychology around these terms accordingly. Indeed, if you expect folk psychology to supply possession conditions for particular mental state concepts as individuated by everyday notions, you constrain not only which concepts can appear in your theory but also the relation in which individuals are required to stand to them. Insofar as these are aims of philosophical accounts of

folk psychology, proposed accounts are linked inextricably to notions we are consciously aware of.

In contrast with this kind of approach, I have argued that the assumption that the theoretical concepts called upon to account for human mindreading abilities are all available introspectively, is unwarranted. A main aim is to establish what are the appropriate theoretical concepts and principles that account for mindreading, not to answer questions about the definition of 'belief'. Maybe you don't even *get* the semantics of reflective commonsense terms like 'belief' from I-psychology, any more than you get the semantics of reflective commonsense terms like 'word' from I-language. Such a scenario is neither implausible nor particularly worrying from the perspective pursued here. However, it could be considered a serious shortcoming for a different kind of account. Therefore, it is important to recognize what is at issue for alternative accounts of some phenomenon before they can be seen as competing.

The situation is not dissimilar to discussions surrounding scientific approaches to language and purported evaluations of them. Someone might contend, for example, that adopting a Chomskyan approach to language leads to having to say that there are no public languages, only idiolects, and that that is clearly not true. One answer to this concern is to accept that this is indeed the correct conclusion, but point out that the only sense in which the claim that there are no public languages was supposed to hold is when language is understood as I-language, and so the objection is misplaced as it can be resolved by refining terminology. However, a less spurious concern that could still be raised might be cast like this: *there is no way to address* questions about public languages and their properties from within a Chomskyan perspective. And that for some people is clearly a problem, if what they are interested in when it comes to language is precisely that sort of issue rather than explaining acquisition or capturing grammaticality judgements (or the goals of naturalistic inquiry more generally).

Bearing this in mind, it is clearly useful to be aware of the aspects of the pre-theoretic concept *language* people from different intellectual backgrounds feel ought

to be addressed, and which of these are left out of a Chomskyan account. Similarly it would be good to be clear which are the bits of the notion of *folk psychology* people would most like to see explained and for what reasons, as it is likely that many concerns will not be addressed from the perspective advocated here. However, even if a naturalistic account along these lines has limitations, in its defence is the fact that it is properly scientific. Furthermore, it was conceded from the outset that we probably get more insight into people when it comes to certain kinds of questions about language and psychology by reading novels and watching films than we do from science, or indeed philosophy.

So, maybe part of the problem is that when it comes to folk psychology (mindreading), what people expect is precisely an account of their '*folk* concepts' in the sense of their everyday concepts about psychology (vernacular commonsense in my terminology). Whereas naturalistic inquiry is interested primarily in an account of the concepts of commonsense psychology, whatever those are.¹¹² I have argued that our everyday concepts about minds (e.g. *belief*) are only suggestive of what provides our commonsense understanding of psychology, just as everyday concepts about language (e.g. *word*) are only suggestive of what fully explains linguistic ability. The everyday concepts are not completely wrong. There are rough counterparts to the concept of *word* posited by linguistic theory (e.g. *lexeme*, *head*). There may be rough counterparts to the concept of *belief* in I-psychology (e.g. *judging*). We might say that the everyday concepts are indicative of categories that feature in linguistic and folk psychological understanding respectively. Indeed, if common sense tells you how to chunk up the world, it makes sense that the lexical resources of natural language should reflect these divisions and not others. Concepts of common sense may be closer to the resources of natural language than are scientific categories. "[T]he

¹¹² Phrases like 'the folk concept of x' or 'the commonsense concept of x' are ambiguous between everyday notions and the theoretical constructs of core common sense. Once the distinction is recognised it is usually clear which is at issue, but it pays to be aware of it.

concepts expressed and distinctions developed in normal language use give us insight into the patterns of thought and the world of ‘common-sense’ constructed by the human mind” (Chomsky 1975, p. 4).

However, vernacular commonsense concepts may be both too wide and too narrow in scope to be helpful on their own in recovering core commonsense principles. They are too wide because, geared as they are towards shifting human interests and concerns, it is unlikely that there will be any one thing in the theory (of I-language or I-psychology) with all of the properties associated pre-theoretically with e.g. *word*. And they are too narrow because, as we saw from consideration of the platitudes listed above, they underspecify what is needed, and appealing to them we underestimate the complexity of the knowledge structures called upon to deliver certain abilities. Talk of *beliefs* not only carries connotations superfluous to the theory, a problem avoided to some extent by introducing a theoretical term like *judging*, but also suggests that there is one uniform category of belief in operation when in fact there may be a variety of them (for instance, different sorts of JUDGING depending on, e.g., strength of assumption, or reliability of information source, or origin of representation, etc.).¹¹³

Having analysed and defused the potential for disanalogies between language and commonsense psychology, we can now return to comparing the two domains. To sum up where we are at the moment: I have drawn a distinction between knowledge of language and knowledge of mindreading on the one hand, and reflective concepts related to mindreading and reflective concepts related to language on the other. As we have seen, much interest in folk psychology has focussed on reflective concepts related to mindreading, an area closer to some kind of ‘folk grammar’ than to I-language. In the next section I look at reflective commonsense notions of language, what they are and how they arise.

¹¹³ See also Sperber (1997) on a distinction between intuitive and reflective belief.

5.4 LANGUAGE ACCORDING TO REFLECTIVE COMMON SENSE

Commonsense understanding is usually divided into areas that roughly correspond to major areas of scientific investigation, hence interest in commonsense *physics*, commonsense *biology*, commonsense *psychology*.¹¹⁴ Of course, different areas of commonsense understanding might not match directly with traditional boundaries between sciences (indeed boundaries between sciences tend not to be strictly ‘enforced’), but let’s assume that there is a rough correspondence. It might then be asked why the focus on these commonsense domains and not others, such as perhaps commonsense *chemistry* or commonsense *astronomy*? The answer is that an important factor motivating the investigation into commonsense understanding is the search for areas where there is a specific mental structure – or learning theory (LT) – that accounts for the abilities in question and explains their acquisition; but in the areas of chemistry and astronomy humans just don’t seem to have the level of systematic expectations about the world that they do about physical or biological matters. To be sure, they can construct (non-scientific) theories about chemical or astronomical aspects of their experience (about combustion or constellations, say), but they do not appear to do so automatically and with the ease associated with more familiar commonsense domains (see Vosniadou 1994). Similarly, humans have folk theories about, say, gardening or medicine, but again these systems of understanding do not appear to be likely candidates for universally arising cognitive abilities, but rather specializations of individuals or groups with particular interests.

Now, what about commonsense ideas about language? Do they amount to a ‘folk’ linguistics on a par with commonsense physics and commonsense biology, or

¹¹⁴ As noted above, we need to distinguish between core common sense and reflective common sense contributions to the commonsense view of some area. Abilities in these domains (physics, biology, psychology) can be due to innate endowment, depend on development and specific experience, or result from introspective re-elaboration and reflection.

are they more arbitrary (in terms of what they are and who has them), like folk astronomy or chemistry or folk medicine or gardening?

The current question, then, is whether salient features of products of a language faculty are the kind of phenomena humans develop stable concepts of, as well as possessing an I-language. And if so, whether this is because some part of commonsense understanding is directed back onto the output of a language faculty. I will argue that there are such concepts, and that they arise in the same way as reflective commonsense notions arise for other domains where humans are equipped with an LT. The content of knowledge of language is the subject of existing research programmes in linguistics. The idea of reflective commonsense knowledge about language is less discussed. Let us start by compiling a list to show what I have in mind. We will see that we can pull together quite readily a wide array of familiar concepts about or associated with language according to common sense, and that these are very easy to categorise. Here is an initial cross-section, followed by a brief discussion. Each category is provisionally labelled:

- concepts like word, utterance, clause, pause, and other *perceptually* salient features of language;
- concepts like intonation, alliteration, rhyme, and other *phonologically* salient features of language;
- concepts like noun, verb, adjective, adverb, preposition, and other *syntactically* salient features of language;
- concepts like subject, object, agent, and other *thematically* salient features of language;
- concepts like proper name, common name, pronoun, and other *referentially* salient features of language
- concepts like meaning, reference, ambiguity, synonymy, homonymy, and other *semantically* salient features of language;
- concepts like sarcasm, irony, metaphor, and other *rhetorically* salient features of language;
- concepts like communication, miscommunication, saying, implicating, and other *communicatively* salient features of language;

- concepts like asserting, denying, promising, questioning, lying and other *performatively* salient concepts of language.

Words can be preceded and followed in speech by brief pauses and in writing by spaces; they can appear in isolation and are units of meaning. Words are thus salient in terms of both external (world) and internal (mental) environments. Alliteration and rhyme specifically exploit particular phonological properties of words. But words are not categorised just according to their sounds and meanings. It can be recognised that different sorts of words have different functions – words that name things, words that describe things, words that label actions – and, once this has been noted, it is quite easy also to taxonomise words accordingly. Given lists of words from different parts of speech, children can readily group them together with minimum instruction. Thematic roles are also familiar, and perhaps are represented not only as part of syntax but as conceptual knowledge too.¹¹⁵ People are well aware that words are used as labels and that intuitively there is a difference between proper names and common nouns, and they are aware of the referential properties of names and pronouns. That you can refer to the same thing in different ways and that words can have more than one meaning is part of most people's reflective knowledge. Irony and metaphor arise and are interpreted naturally and do not need to be specifically taught. Ideas about communication are particularly important to the commonsense conception of language, for language is inextricably linked to communication according to common sense. Related to the idea that language is above all for communicating, there is widespread recognition that you can do other things with words as well. The communicatively and performatively salient notions about language begin to overlap with folk psychological ideas about intentions.

We can make some immediate observations. First, the suggestion is not that the terms on the list have to be learnt for the concepts they express to be recognised – one can notice, say, repetition of similar sounds without necessarily having labelled a category. That children's nursery rhymes and stories exploit sound patterns much to children's delight suggests that children are appreciating alliteration even if they don't

¹¹⁵ See Chomsky 1980, p. 54-55. Some syntactic features have semantic import that may relate to concepts present already as part of commonsense understanding – so, for example, a syntactic feature [+agent] maps onto a commonsense concept of *agent*.